THE UNIVERSITY
of EDINBURGH

The Epigenetics of Paediatric Crohn's Disease

THE UNIVERSITY
of EDINBURGH

Alex Adams

A thesis submitted for the degree of Doctor of Philosophy

2016

Contents

List of Figures

List of Tables

Declaration

I declare that this thesis has been composed solely by myself and that it has not been submitted, in whole or in part, for another degree or qualification.

Some sections of chapters 3, 4, 5, & 6 have been published in the following article, of which I was the joint first author, drafted the manuscript, prepared the figures, performed the majority of the experimental work and performed the analysis, except initial processing and cell type correction of the Illumina 450k data which was performed in collaboration with Dr Nick Kennedy.

> Adams AT, *et al.* Two-stage Genome-wide Methylation Profiling in Childhood-onset Crohn's Disease Implicates Epigenetic Alterations at the VMP1/MIR21 and HLA Loci. *Inflammatory Bowel Diseases.* 20(10): 1784–93. 2014.

Some sections of chapter 2, in particular sections 2.4 & 2.5 are planned to be published in a review article, of which I am the first author and drafted the manuscript.

Some sections of chapters 3 and 6 have appeared in patent applications [3, 5].

Where work was performed in collaboration with colleagues, or datasets which were generated by others in our group have been used, this is clearly stated in the text (see page xvii).

Alex Adams
1st August, 2016

Notes

All genomic coordinates are for the GRCh37/Hg19 assembly unless stated otherwise.

P values are uncorrected unless stated otherwise and are not normally calculated beyond 2.2×10^{-308}.

This document is available digitally, with all internal references and citations functioning as hyperlinks.

With the exception of proprietary software necessary for running the pyrosequencing and qPCR machines this work was entirely produced using free and open source software.

Word count: 50309 (TeXcount)

Acknowledgements

This work was primarily funded by CICRA[1], with additional funding from the CSO[2] and IBD Character[3].

I would like to thank Prof. Jack Satsangi and Dr Elaine Nimmo for their guidance and supervision, Dr Nick Kennedy for extensive technical advice, and Dr Nidhi D. Sharma & Ms Kate O'Leary for everything.

Additionally I would like to thank the entire gastrointestinal research group in the Western General Hospital, our collaborators in the BISCUIT, PICTS, and IBD Character studies, and all the patients and volunteers whose gave the blood samples and biopsies this work is based on.

I received extremely generous support from the RMBF[4] for which I am very be grateful.

Finally, I would not have been able to undertake this work without the support of my family and thank them for everything they have done.

[1]Crohn's in Childhood Research Association, registered charity 278212
[2]Chief Scientist Office, Scottish Government
[3]www.ibdcharacter.eu
[4]Royal Medical Benevolent Fund, Registered charity No 207275

Abbreviations

Contributions

ATA	Alex Adams
NAK	Nick Kennedy
ERN	Elaine Nimmo
CLN	Colin Noble
KRO	Kate O'Leary
RK	Rahul Kalla
NTV	Nick Ventham
WTCRF	Wellcome Trust Clinical Research Facility Western General Hospital Edinburgh

General

5-ASA	5-Aminosalicylic Acid
5caC	5-Carboxylcytosine
5fC	5-formylcytosine
5hmC	5-Hydroxymethylcytosine
5mC	5-Methylcytosine
6MP	6-Mercaptopurine
aa	Amino Acid
AIEC	Adherent Invasive *Escherichia coli*
ATP	Adenosine triphosphate
AUC	Area Under the Curve
AZA	Azathioprine
BISCUIT	Bacteria in IBD in Scottish Children Under-going Investigation before Treatment study [244]
CARD	Caspase (Activation and) Recruitment Domain
CD	Crohn's Disease
CGI	CpG Island
CHD	Coronary Heart Disease
ChIP-Seq	Chromatin Immunoprecipitation Sequencing

CI	Confidence Interval
CpG	Cytosine-phosphate-Guanine dinucleotide
CRC	Colorectal Cancer
CRP	C-Reactive Protein
DHS	DNase I Hypersensitivity Site
DMEM	Dulbecco's Modified Eagle Medium
DMR	Differentially Methylated Region
DNMT	DNA Methyltransferase
dNTP	Deoxynucleoside triphosphate
DSS	Dextran Sulfate Sodium
EDTA	Ethylenediaminetetraacetic acid
ER	Endoplasmic Reticulum
EWAS	Epigenome-Wide Association Study
FCS	Foetal Calf Serum
FDR	False Discovery Rate (Benjamini-Hochberg)
FWER	Familywise Error Rate
GCC	Glucocorticoid
GO	Gene Ontology
GWAS	Genome-Wide Association Study
H#K#	Histone# Lysine#
HAT	Histone Acetylase
HBSS	Hank's Balanced Salt Solution
HC	Healthy Control
HDAC(i)	Histone Deacetylase (inhibitor)
HDL	High Density Lipoprotein
HR	Hazard Ratio
IBD	Inflammatory Bowel Disease
IBDU	Inflammatory Bowel Disease Unclassified
IBS	Irritable Bowel Syndrome
IEC	Intestinal Epithelial Cell
IFN	Interferon
IHC	Immunohistochemistry
IκB	Inhibitor of (Nuclear Factor)κB
IKK	IκB Kinase
Illumina 27k	Illumina HumanMethylation27 BeadChip
Illumina 450k	Illumina HumanMethylation450 BeadChip
IQR	Interquartile range
i.v.	intravenous
KO	Knockout
LD	Linkage Disequilibrium
LDA	Linear Discriminant Analysis
LDL	Low Density Lipoprotein
LMR	Low-Methylated Region
LPS	Lipopolysaccharide
LRR	Leucine Rich Repeat

MAF	Minor Allele Frequency
mir	MicroRNA
MDP	Muramyl Dipeptide
MTX	Methotrexate
NBD	Nucleotide Binding Domain
NF$\varkappa$B	Nuclear Factor $\varkappa$B
NGS	Next Generation Sequencing
NK	Natural Killer cells
NLR	NOD-Like Receptor NBD and LRR containing
nt	Nucleotide
OCP	Oral Contraceptive Pill
OR	Odds Ratio
PBL	Peripheral Blood Lymphocytes
PBMC	Peripheral Blood Mononuclear Cell
PDF	Probability density function
PI3K	Phosphatidylinositol 3-Kinase
PI3P	Phosphatidylinositol 3-Phosphate
PICTS	Paediatric-onset IBD Cohort and Treatment Study [653]
PSC	Primary Sclerosing Cholangitis
RA	Rheumatoid Arthritis
ROS	Reactive Oxygen Species
SAM	S-Adenosyl Methionine
SCFA	Short Chain Fatty Acid
SLE	Systemic Lupus Erythematosus
SNP	Single Nucleotide Polymorphism
T1D	Type 1 Diabetes
TF	Transcription Factor
T_H	T Helper Cell
TLR	Toll-Like Receptor
TNBS	2,4,6-Trinitrobenzenesulfonic acid
TNF	Tumour Necrosis Factor
TPN	Total Parenteral Nutrition
TSS	Transcription Start Site
UC	Ulcerative Colitis
UC-CRC	UC-associated colorectal cancer
UPR	Unfolded Protein Response
UTR	Untranslated Region
VTE	Venous Thromboembolism

Abstract

Inflammatory bowel disease (IBD), comprising Crohn's disease (CD) and ulcerative colitis (UC), is a complex disease with multiple aetiological factors including genetics, environmental exposures and microbiota. Epigenetic alterations such as DNA methylation, histone modification, and microRNA activity may indicate or facilitate the interactions between genetic and non-genetic processes.

This thesis describes a study of genome-wide changes in DNA methylation between children with Crohn's disease and symptomatic controls performed with the Illumina HumanMethylation450 platform. At the time of publication [4] this work was the largest epigenome-wide study of IBD, and described 65 individual CpGs (Bonferroni corrected) and 19 differentially methylated regions where there were significant CD-associated methylation changes.

A highly significant ($p=1.97\times10^{-15}$) hypomethylated differentially methylated region (DMR) in the penultimate exon of the critical autophagy gene *VMP1*, appearing to correspond to the microRNA mir21, has been replicated in 3 adult cohorts by pyrosequencing, and found in 2 subsequent Illumina 450k adult epigenome-wide studies.

Hypomethylation of a highly significant CpG ($p=4.47\times10^{-15}$) in *RPS6KA2* has also been replicated in 3 adult cohorts by pyrosequencing and 2 adult cohorts with the Illumina 450k platform.

In line with our previously published work [458] methylation findings were demonstrated to be enriched in proximity to genetic risk loci, with an improved analytical technique correcting for Illumina 450k probe density and hidden confounding factors.

Section 3.3.7 demonstrates enrichment of a DNA motif matching the binding sites for the IBD-associated transcription factors GLI1 and RUNX3 at significant methylation changes.

Examining methylation at any combination of two significant CD-associated differentially methylated positions could distinguish between children with CD and paediatric symptomatic controls requiring colonoscopy to rule out IBD. These two probe biomarkers were tested in a separate paediatric cohort with AUCs of 0.79–0.98 (median 0.93)

These data confirm the existence of significant DNA methylation differences between children with Crohn's disease and symptomatic controls, offering potential insight into established genetic risk loci, and novel risk genes and processes.

Characterisation of the observed DNA methylation changes and power calculations based on experimental results rather than simulated ones, will be of use to anyone planning a similar study in a complex disease in a heterogenous cellular population.

Lay Summary

Crohn's disease is a chronic disease that causes severe inflammation of the bowel and sometimes also affects other organs. It is closely related to another disease, ulcerative colitis, and together they are called inflammatory bowel disease (IBD). These diseases may start at any point in life, and although treatments exist to improve symptoms and reduce inflammation, there is no cure, and many patients eventually require operations to remove large portions of the bowel. The best understanding of the cause of Crohn's disease is that the body's immune system reacts abnormally to the bacteria which live in the bowel, with certain inherited factors and environmental exposures also being partly responsible.

Epigenetics (literally 'in addition to genetics') is the field of science that deals with all mechanisms which affect the level of activity of genes, without changing the sequence of bases that make up the gene. Epigenetic modifications help us explain how the same genome can produce many different cell types, how people with the same genomes (identical twins) may not both suffer from a disease, how our environment, medications and smoking habits may affect our disease risk, and part of how genes are turned on and off throughout development. Recent results have hinted that differences in epigenetic regulation of genes may be involved in the development of complex diseases such as IBD, and others including schizophrenia, asthma, and type 1 diabetes.

The epigenetic modification studied here is DNA methylation. DNA is made up of a sequence of four chemicals joined together, there are called bases or nucleotides and are commonly shown as A,C,G, and T. When a C, (cytosine) is immediately followed by a G (guanine) this is called a CpG dinucleotide. The C of a CpG is sometimes modified by the attachment of an extra chemical called a methyl group with the resulting base being called methylcytosine (5mC). The modification of cytosine has an effect on the accessibility of the surrounding genetic sequence to the DNA-reading machinery of the cell, and the function of special short genetic sequences such as promoters, enhancers, and silencers,

which are important in fine-tuning the activity of genes.

Using a process called bisulfite conversion, which affects C but not 5mC allows methods to read DNA to distinguish between the two forms and find the proportion of each which is present for any CpG. We used Illumina 450k chips to measure the proportions of C and 5mC at over 450,000 CpGs throughout the genome, and see if any CpGs have different proportions of methylation in people with Crohn's disease compared to healthy people.

We investigated DNA methylation in 36 children with Crohn's disease and 36 children who had symptoms requiring investigation but were found to be healthy and did not go on to develop any bowel diseases. As epigenetic markers are partly dependant on age and sex we matched these as closely as possible between the cases and controls. Even when inflammatory bowel disease is fairly inactive it changes the proportions of the different cells in blood and so we used statistical techniques to control for this effect.

The analysis found 65 areas where there was a significant difference in methylation related to the presence of Crohn's disease. Many of the genes where these differences were found are known to be involved in processes related to the development of Crohn's disease, or are in regions of the genome where we know there are some inheritable risk factors for Crohn's disease.

We have measured methylation in many of the strongest findings in separate groups of adult patients by the same technique and another technique called pyrosequencing, thus indicating that these are real findings rather than anomalies.

For two of the strongest results we have also gone on to look for differences in methylation in other diseases, and to see if we can find any effects the differences in methylation causes to the levels or activity of the genes in question.

The control children had symptoms suggestive enough of Crohn's disease that their doctors performed an endoscopy to investigate — an uncomfortable and invasive procedure. We have shown that a blood test based on the methylation readings at any two of the 65 strongest results are sufficient to predict the presence of Crohn's disease in a group of children like these with an accuracy of up to 95%.

Chapter 1

Introduction

Signs, symptoms & natural history Inflammatory bowel disease (IBD) refers to the separate but related conditions Crohn's disease (CD) and ulcerative colitis (UC). Additionally there is a significant minority of patients in whom the exact diagnosis is unclear (IBD unclassified – IBDU, previously indeterminate colitis – IC), most of whom later have a clear diagnosis of CD or UC [716], but some remain unclassified [632].

History and examination along with inflammatory markers such as C-reactive protein and faecal calprotectin can be highly suggestive of IBD, but the diagnosis is generally made on the macroscopic and microscopic appearance of the intestine at endoscopy. Crohn's disease classically causes a patchy transmural inflammation, most commonly affecting the ileum and colon, but can manifest anywhere in the gastrointestinal tract, whereas ulcerative colitis generally causes a diffuse continuous inflammation extending from the rectum. A comparison of the features of Crohn's disease and ulcerative colitis is shown in table 1.1.

Crohn's disease is classified by age of onset, disease location and behaviour, with young age of onset, extensive disease and stricturing or penetrating behaviour associated with worse outcomes [573, 370] (detailed in table 2.5, page 45).

The most common symptoms of Crohn's disease are abdominal pain and diarrhoea (bloody diarrhoea being more strongly correlated with ulcerative colitis) accompanied by general malaise, fever, anorexia and weight loss [445]. Abscesses, strictures, fistulae, and peri-anal disease may be present at diagnosis, and become more prevalent over time [573]. Extraintestinal manifestations of Crohn's disease are common. Additionally, the consequences of chronic inflammation, malabsorption of vitamins and nutrients, long-term immunosuppressive treatment, and

shared risks for other autoimmune and inflammatory conditions lead to further disability (see table 1.2).

The natural history is of a relapsing and remitting inflammatory state, with periods of remission where patients can be quite well. However, despite reductions in mortality in some areas [294, 664] long-term mortality remains elevated in IBD (hazard ratio (HR): 1.1 in UC, 1.5 in CD), after falling from an initial high risk in the year following diagnosis (HR: 2.43 in UC, 3.69 in CD) [294, 60]. IBD places a significant burden on healthcare systems and economies due to the resultant morbidity and mortality, requirement for costly ongoing medical treatments, recurrent hospitalisations and often repeated surgical intervention, and these factors also place a considerable psychosocial burden on patients [100].

Treatment strategies: medical and surgical The cornerstone of treatment for Crohn's disease and ulcerative colitis is medication to control the inflammatory response and induce remission, with ongoing medication to maintain remission, and surgical resection when required. General immunosuppressants include glucocorticoids (e.g. prednisolone, methylprednisolone or i.v. hydrocortisone), methotrexate, 5-aminosalicylic acid (mesalazine) and azathioprine or mercaptopurine. In the past 10 years, more targeted biologic therapies against TNFα (e.g. infliximab, adalimumab) or other molecules (Vedolizumab — anti-integrin $\alpha_4\beta_7$) have become available [453].

Surgery could be considered curative for ulcerative colitis; for example in a proctocolectomy with an ileal pouch–anal anastomosis only a small cuff of rectal mucosa remains, effectively eliminating the site of inflammation and neoplastic transformation [148]. Despite the effectiveness, safety, high patient satisfaction, and reasonable cost versus biological therapy, less than 30% of ulcerative patients require surgery in their lifetime [445, 148]. In contrast, surgery is required in around 80% of cases of Crohn's disease [659, 445] and is not curative, with most patients experiencing symptomatic and endoscopic recurrence — with heightened risk in smokers, those with penetrating disease and those who have had prior surgery [659].

In very severe cases with intestinal failure, intestinal transplant [208] or multi-organ transplant may be considered. Additionally there is interest in autologous stem cell transplantation [252] and faecal transplantation [123].

Historical perspective Examination of historical sources finds case descriptions and autopsy reports consistent with IBD from the 17$^{\text{th}}$ century [330, 524]

Feature	Crohn's Disease	Ulcerative Colitis
Primary location	Ileum	Colon
Extended location	Mouth to anus	Terminal ileum to rectum
Rectal disease	Rare	Common
Perianal disease	Common	Rare
Distribution	Patchy	Diffuse & continuous
Inflammation	Transmural	Mucosal
Fissures	Deep	Rare & superficial
Abscesses & Fistulae	Common	Rare

Table 1.1: Comparison of Crohn's disease and ulcerative colitis

onwards, and possible cases such as Alfred the Great, King of Wessex (849–899) [130] and King Louis XIII of France (1601–1643) [38]. Ulcerative colitis was a well established condition in the late 19$^{\text{th}}$ century [155], with Burrill Crohn [133] later attributing the first case description to Sir Samuel Wilks in 1859 [695, 696]. Crohn's disease was recognized later, described as chronic interstitial enteritis by Sir Thomas Dalziel in 1913 [141] and regional ileitis by Crohn *et al.* in 1932 [134].

Intestinal tuberculosis is a differential diagnosis for many of the possible historical cases, and indeed the finding that many intestinal granulomas contained neither acid-fast bacilli nor foreign bodies was a pivotal event in recognizing Crohn's disease as a clinical entity [38].

> "The symptoms in all the cases were similar; the characteristic and most striking feature being most violent colic, causing vomiting and occasionally an escape of some blood, and also constant mucus from the bowel [...]
>
> The cases gave the impression that they were probably tuberculous, and yet from the uniform character of the affection it evidently is not so. The affected bowel gives the consistence and smoothness of an eel in a state of rigor mortis, and the glands, though enlarged, are evidently not caseous [...]
>
> As far as I know the disease has not been previously described, but it seems probable that many cases must have been seen and have been diagnosed as tuberculous, and possibly nothing done for their relief.
>
> In regard to treatment, these cases which have come under observation have pursued their course uninfluenced by dietetic or medicinal treatment, and apparently only operation can afford relief, and then only if the disease be limited. Seven out of the nine made a perfect

> recovery after the operation, and one does not hesitate in resecting large portions of the intestine. The subject has been one of great interest to me for some years. My friends the pathologists prefer to call it hyperplastic enteritis, and I can only regret that the etiology of the condition remains in obscurity, but I trust that ere long further consideration will clear up the difficulty."
>
> — T.K. Dalziel, 1913 [141]

Understanding the causes of Crohn's disease Despite over a century of further consideration since the words of Dalziel, much of the the difficulty remains. We have an understanding of some of the processes and pathways involved in the development of IBD: a disruption in the complex homeostasis [47] between the host immune system and the intestinal microbiota in a genetically susceptible individual upon certain environmental exposures leading to an inappropriate inflammatory response. However, many details remain obscured and research into the diagnosis, management and underlying pathology of IBD is of great importance.

Table 1.2: Extraintestinal manifestations of IBD

Condition	Frequency	Risk factors
Musculoskeletal	<35%	
Seronegative arthritis	5–20%	CD>UC
†Type I (Oligoarticular)		Colitis
Type II (Polyarticular)		
Ankylosing Spondylitis	<10%	CD>UC, M>F, HLA-B27
*Clubbing	15% CD, 38% UC	
*Osteoporosis	<15%	Malabsorption, GCC, chronic inflammation, smoking
Dermatological	10–20%	
†Erythema Nodosum	10–15%	CD>UC, M<F
Pyoderma Gangrenosum	<5% UC	CD<UC
Psoriasis	<10%	CD>UC
Aphthous stomatitis	<10%	
†Sweet syndrome	$\ll$ 1%	UC: pustular variant
Hepatobillary		
Primary Sclerosing Cholangitis	<7.5%	CD<UC, HLA-B8, p-ANCA
Cholangiocarcinoma		PSC
Autoimmune hepatitis		CD<UC
Gallstones	<30%	Ileal CD, ileal resection
*NAFLD / NASH	9% UC, 19% CD	GCC, MTX, AZA, TPN
*Pancreatitis	1–1.5%	6MP, AZA, gallstones, duodenal inflammation

Continued overleaf

Table 1.2: Extraintestinal manifestations of IBD

Condition	Frequency	Risk factors
Ocular	2–6%	
†Episcleritis		Colitis
†Scleritis		Colitis
Uveitis		Arthritis, M<F
Renal	4–23%	
Kidney stones	5–15%	Diarrhoea
Secondary amyloidosis	rare	CD>UC, M>F
Haematological		
Thrombosis	HR 2–4	Highly active disease
*Anaemia	19–32%	Blood loss, malabsorption, AZA, 6MP, MTX
Neurological		
Restless legs syndrome	30%	CD
*Vasculitis		Anti-TNF
Stroke / TIA	HR:1.28	Young age, M<F
Sensorineural hearing loss		UC
*Multiple sclerosis		Anti-TNF, CD<UC
*PML		Natalizumab, JC virus
LOW RISK	0.01%	JCV negative
HIGH RISK	1%	JCV positive

Table 1.2: Extraintestinal manifestations and non-gastrointestinal sequalae of inflammatory bowel disease [529, 360, 373, 474, 578, 414, 102, 440]. Frequency estimates and risk factors are poorly defined for some extra-intestinal manifestations due to rarity or, *e.g.* clubbing prevalence estimates date from 1979 [333], and may now be lower due to improved inflammatory control. †Denotes extra-intestinal manifestations following intestinal disease activity, * denotes those partially secondary to drug treatment or chronic illness. GCC: steroids, MTX: methotrexate, AZA: azathioprine, 6MP: 6-mercaptopurine, TPN: total parenteral nutrition, HR: hazard ratio, PML: progressive multifocal leucoencephalopathy

Chapter 2

Literature Review

2.1 Genetic factors & missing heritability

Genetic risk factors Families with multiple cases of ulcerative colitis were identified as early as 1909 but dismissed as coincidence [9]. Later studies found significantly increased risk of IBD in first-degree relatives of IBD patients, with a greater risk where the affected relative has CD (relative risk 13–36 for a sibling with CD, 7–17 for UC), and the highest risk among offspring of two parents with IBD — with an IBD prevalence of 33–52% [168]. While most multiplex families were concordant for which form of IBD they displayed, a quarter of families displayed both, and familial concordance for disease location, behaviour, age of onset, extraintestinal manifestations and number of surgical resections have also been reported [168].

Family history for IBD and autoimmune conditions are more common in patients who develop IBD as children [531, 498]. With decreasing age at which the signs and symptoms of IBD are discovered, increasing proportions of patients have been found to have individually very rare mutations causing disruption of processes such as epithelial barrier function, bacterial clearance, immune cell activity, or cytokine signaling, and as well as findings consistent with IBD may also present other signs of immune system dysfunction. These cases of 'monogenic IBD' are highly variable, though described cases are often associated with a severe disease which responds poorly to conventional treatment [641].

Early genetic studies identified areas of significant linkage, with *NOD2* reported as the first susceptibility gene for Crohn's disease in 2001 [466, 275], and established that the genetic contribution to IBD risk was polygenetic. Genome-wide association studies (GWAS) were undertaken to search for numerous risk

loci with small effects, identifying over 200 genetic risk loci [384, 297, 194, 507] to date (see table 2.1), among the most for any complex disease. The majority of loci are relevant to both CD and UC (68.2%), with more loci for only CD (18.4%) than UC (13.4%) [297, 384]. Loci which were initially reported as correlated with only one disease have shown effects in the other disease with increasingly large GWAS sample sizes and meta-analyses, and further overlaps may yet be reported.

Alongside *NOD2*, other genes which make significant individual contributions to risk such as *IL23R* and *ATG16L1* [297] have been found, and pathways with significant involvement in pathogenesis elucidated. Despite these successes, the total disease variance explained by GWAS loci is small (8.2–13.6% in CD and 4.1–8.2% in UC) [297, 384].

Missing heritability The difference between observed heritability of complex genetic traits and that explained by known genetic risk factors has been dubbed 'missing heritability' [166, 741, 405]. Some of this may be explained by assumptions inherent to the GWAS technique, such as the focus on common variants or common biases such as concentration on European-ancestry populations [405]. As a group, individually rare variants are common but poorly detected by GWAS, as are deletions, duplications, microsatellite repeat expansions, translocations and inversions [405, 166]. Therefore sequencing common SNPs overly simplifies the genetic landscape.

Epistatic and parent-of-origin effects may be underestimated [166], and the sample sizes needed to characterise epistatic interactions are substantially larger than the numbers needed to detect susceptibility loci by GWAS [741].

It may be that the total heritability is overestimated due to epistatic interactions, shared familial environmental exposures, gene-environment interactions, or a small subset of cases being due to undetected Mendelian inheritance with incomplete penetrance and that the missing heritability is in fact phantom heritability [741, 405, 101]. There is also evidence that the heritability of Crohn's disease was overestimated by early twin concordance studies [235].

It has been posited that epigenetic mechanisms may be responsible for some portion of true missing heritability [633]; however, the importance of transgenerational epigenetic inheritance to human disease is not well understood and is somewhat controversial.

Even if the true heritability of Crohn's disease is not accurately known, it is clear that non-genetic effects are also important, and again epigenetics may have some role.

Reference	Diagnosis		Case:Control		Replication
Yamazaki, 2005 [710]	CD	$\mathbb{J}$	484:752	$\mathbb{E}$	363:372, ⌂ 580
Duerr, 2006 [158]	IBD	$\mathbb{E}$	547:548	✡	401:433, ⌂ 883
Hampe, 2007 [239]	CD	$\mathbb{E}$	735:368		498:1032, ⌂ 380
Libioulle, 2007 [377]	CD	$\mathbb{E}$	537:913		1266:559, ⌂ 428
WTCCC, 2007 [693]	CD	$\mathbb{E}$	1748:2872		
Parkes, 2007 [486]	CD	$\mathbb{E}$	1182:2024		
Rioux, 2007 [521]	iCD	$\mathbb{A}$	946:977		353:207, ⌂ 530
Franke, 2007 [193]	CD	$\mathbb{E}$	393:399		942:1082, ⌂ 375
Raelson, 2007 [507]	CD	$\mathbb{A}^{\mathrm{Q}}$	⌂ 477	$\mathbb{E}^{\mathrm{D}}$	750:828, ⌂ 521
Franke, 2008 [191]	UC	$\mathbb{E}$	1167:777		1855:3091
Barrett, 2008 [41]	CD	$\mathbb{M}$	3230:4829		
Kugathasan, 2008 [349]	pIBD	$\mathbb{A}$	1011:4250		173:3481
UK-IBD & WTCCC, 2009 [642]	UC	$\mathbb{E}$	2361:5417		2321:4818
Asano, 2009 [27]	UC	$\mathbb{J}$	752:2031		632:1026
Silverberg, 2009 [572]	UC	$\mathbb{E}$	1052:2571	$\mathbb{A}$	769:727, $\mathbb{E}^{\mathrm{I}}$ 633:415
Imielinski, 2009 [284]	pIBD	$\mathbb{A}$	2784:7315		482:1696, $\mathbb{AE}$ 531:4109
Haritunians, 2011 [247]	MR-UC	$\mathbb{A}$	324MR, 537UC, 2601HC		
Franke, 2010 [192]	CD	$\mathbb{M}$ $\mathbb{E}$	6333:15056		15694:14026
McGovern, 2010 [422]	UC	$\mathbb{E}$	2693:6791		2009:1580
McGovern, 2010 [423]	CD	$\mathbb{E}$	896:3204		1174:357
Torkvist, 2010 [628]	IBD	$\mathbb{E}^{\mathrm{S}}$	736CD, 935UC, 1460HC		

Reference	Diagnosis		Case:Control		Replication
Franke, 2010 [194]	UC	$\mathbb{E}^{D}$	1043:1703	$\mathbb{E}$	2539:5428
Anderson, 2011 [16]	UC	$\mathbb{M}$ $\mathbb{E}$	6687:19718		9628:12917
Okada, 2011 [468]	IBD	$\mathbb{J}$	979CD, 748UC, 905HC		
Umeno, 2011 [643]	IBD	$\mathbb{M}$	29606:47210		
Kenny, 2012 [314]	CD	✡^{A}	907:2345		971:2124
Julia, 2012 [299]	CD	$\mathbb{E}^{E}$	1341:1518		1365:1396
Huang, 2012 [274]	CD		*Reanalysis of WTCCC, 2007 [693]*		
Jostins, 2012 [297]	IBD	$\mathbb{M}$	>75000		
Yamazaki, 2013 [711]	CD	$\mathbb{J}$	372:3389		1151:15800
Yang, 2013 [713]	UC	$\mathbb{K}$	388:739		417:732
Yang, 2014 [714]	CD	$\mathbb{K}$	533:800		521:732, 1258:977
Liu, 2015 [384]	IBD	$\mathbb{M}$	42950:53536		
Huang, 2015 [273]	IBD	$\mathbb{N}^{A}$	1511:1797		

Table 2.1: Published IBD GWAS studies. Numbers shown after QC. UK-IBD: UK IBD Genetics Consortium, WTCCC: Wellcome Trust Case Control Consortium.
iCD: Ileal CD, pIBD: paediatric IBD, MR-UC: medically refractory UC.
$\mathbb{A}$merican: USA primarily or wholly of European ancestry, ^{A}: African American, ^{Q}: Québécois,
$\mathbb{E}$uropean, ^{D}: German, ^{E}: Spanish, ^{I}: Italian, ^{S}: Swedish
$\text{✡}^{(A)}$ Jewish (Ashkenazi), $\mathbb{J}$apenese, $\mathbb{K}$orean, $\mathbb{M}$eta-analysis
⌂ Trio of proband and parents, (number of family trios, not number of individuals)
population of replication cohorts match that of primary cohorts unless stated

2.2 Core pathways

2.2.1 NOD2

The first Crohn's disease associated gene The IBD1 locus on chromosome 16 was identified as a Crohn's disease susceptibility locus in 1996 [276] with multiple independent replications in the following years [138, 75, 20, 617]. *NOD2* was identified as the responsible gene when several CD-associated SNPs within *NOD2* were identified in 2001 [466, 275, 238].

NOD2 (previously known as CARD15) is one of 22 members of the NLR family ('nucleotide-binding domain and leucine-rich repeat containing', or commonly 'NOD-like receptor') [624]. In addition to the nucleotide-binding domain (NBD or NOD domain) and leucine-rich repeats (LRR) implied by their name, NLR family members contain an N-terminal effector domain; either an acidic transactivation domain, pyrin domain, caspase recruitment domain (CARD) or Baculoviral inhibitory repeat (BIR)–like domain. In the case of NOD2, it has 2 N-terminal CARD domains [624]. Similar LRR domains are found in Toll-like receptors (TLR) which recognize a variety of triggers to the innate immune system.

Muramyl dipeptide NOD2 is expressed in many cell types [726] including monocytes [467] and Paneth cells [733] and responds to the presence of muramyl dipeptide (MDP) [286], a component of gram-negative and gram-positive bacterial cell walls. MDP can be derived from intracellular bacteria, injected into the cell by bacterial secretion systems (e.g. *Helicobacter pylori* [666]), transported into the cell by peptide transporters (including SLC15A1, 2, 3 & 4) or by endocytosis from peptidoglycan fragments produced by bacterial cell division [73, 491, 490].

Structure & function The C-terminal LRR of NOD2 is generally thought to be the site of interaction with MDP [607], although direct interaction with the NBD has been proposed [432]. Binding of adenosine triphosphate (ATP) to the NBD allows interaction with MDP and subsequently homo-oligomerisation of NOD2 facilitated by CARD-CARD and NBD-NBD interactions [432]. In addition to forming homo-oligomers, NOD2 directly interacts with a large number of proteins [73] including autophagy proteins (ATG16L1 [631] and BCN1 [298]), other NLR [674, 73] (NLRC4, NLRP1, NLRP3 and NLRP12) and others includ-

ing SLC15A3/4 [452] CARD8 [672], CARD9 [272], HSP90 [363], SOCS3 [363], TLE1 [459] and RIPK2, many of which have CD-associated genetic variants[1].

NFkB signalling NOD2 interacts with the serine-threonine kinase RIPK2 by a CARD-CARD interaction, which in turn leads to the activation of NFκB via the IκB Kinase (IKK) pathway [467]. Signalling is turned off by hydrolysis of the NBD-bound ATP by NOD2 and dissociation of the oligomer components [432]. To prevent excessive NOD2 stimulation, MDP causes decreased expression and increased proteosomal degradation of NOD2 and RIPK2, dissociation of the chaperone protein Hsp90, and induction of SOCS3 expression which further degrades NOD2 [363]. NOD2 has a basal level of NFκB signalling, which is significantly increased with mutations causing Blau syndrome (page 12) and decreased with CD-associated mutations, and may represent autoinhibition of the LRR by the CARD domains when unstimulated [467, 607, 700].

NFκB is a transcription factor with central roles in inflammation such as regulating the production of inflammatory cytokines (TNFα, IL-1β [254], IL-6, IL-8) and defensins [606]. It is also involved in cell activation and proliferation in inflammation and cancer [629] and links inflammation to metabolic pathways [629] and cancer [152].

MAPK signalling is also induced by NOD2 signalling in a largely IL-1β dependant mechanism [254] and is important in NOD2-induced autophagy [73] and cytokine release [254].

Genetic variants in NOD2 CD-associated NOD2 variants are less responsive to MDP [286, 254], do not cause autophagy in response to MDP [432], have a reduced NFκB response, and impaired recruitment of ATG16L1 to the plasma membrane [631].

Genetic variants in *NOD2* have also been associated with Blau syndrome, Early onset sarcoidosis [700] and NAID (NOD2-associated autoinflammatory disease) [718, 717]. Blau syndrome and Early onset sarcoidosis are now recognized as the autosomal dominant and spontaneous forms of the same childhood-onset disease [528]. There is an overlapping, but distinct phenotype in NAID, which occurs in adults [718]. Blau syndrome classically includes a triad of granulomatous dermatitis, polyarticular boggy arthritis and uveitis, but may also include erythema nodosum and granulomatous disease of various organs, which are his-

[1] *ATG16L1* (see table 2.2), *CARD8* [421], *CARD9* [735, 297], *NLRP1* [137], *NLRP3* [137], *TLE1* [459], *RIPK2* [297].

tologically distinct from granulomas seen in Crohn's disease [700]. NAID causes intermittent dermatitis and fever, arthritis, dry mouth and eyes (but no uveitis), and abdominal pain and non-bloody diarrhoea with no evidence of IBD [717]. Blau syndrome genetic variants are mostly found in the nucleotide-binding domain and cause overactive NOD2 [607, 700] whereas NAID has some overlap with CD-associated variants such as R702W, although IVS8^{+158} predominates [717].

The paradoxical finding of gain-of-function and loss-of-function NOD2 variants both leading to granulomatous inflammatory diseases has yet to be fully explained [300]. It may be that in IBD a primary immunodeficiency in macrophages leads to poor clearance of bacteria and a resultant T cell mediated inflammation [101, 581]. Additionally pathways other than NF$\varkappa$B signalling such as IL-10 production [463], antigen presentation and autophagy may be important.

Other downstream effects NOD2 activation leads to inducible nitric oxide synthase (iNOS) expression and activation of the Sonic Hedgehog (SHH) signalling pathway [210]. NOD2 activation also leads to recruitment of NOD2 and ATG16L1 to bacterial entry sites in the plasma membrane and induces autophagy in a RIPK2 and NF$\varkappa$B independent mechanism [631]. The identification of CD-associated genetic variants in *ATG16L1* and other autophagy related genes has led to increasing interest in the role of autophagy in Crohn's disease.

2.2.2 Autophagy

Autophagy is a highly conserved process triggered by cellular stressors such as microorganisms, nutrient starvation, ER stress, hypoxia and tumours. There are three types of autophagy; chaperone-mediated autophagy (CMA), microautophagy and macroautophagy. CMA is a process in which specific soluble cytosolic proteins are transported across the lysosomal membrane and degraded [472]. In microautophagy cytosolic proteins are included within an invagination of the lysosomal membrane from which a vesicle is formed and then degraded by the lysosomal hydrolases [375].

In macroautophagy (commonly referred to as autophagy) substrates are first enveloped by a phagophore membrane, which then closes to form an autophagosome before merging with a lysosome to form an autolysosome [182, 372]. Selective processes for the autophagy of mitochondria, ribosomes, peroxisomes and endoplasmic reticulum have been described, as well as a process for the digestion of microorganisms - xenophagy [372, 511].

Autophagy allows recycling of cytosolic proteins and organelles to maintain critical pathways such as the tricarboxylic acid (TCA) cycle in response to nutrient deprivation and to support cells during more quiescent periods when anabolic processes are not being stimulated by growth factors [399]. Autophagy also functions to remove organelles due to damage, age, or overabundance [399].

Normal autophagy is important to development, non-apoptotic cell death, and the innate and adaptive immune systems (MHC class II antigen presentation). Altered autophagy function is frequently described in association with tumours, ageing and degenerative conditions, autoimmune conditions and inflammatory diseases including Crohn's disease [372, 257]. Autophagy genes linked to Crohn's disease are shown in table 2.2.

The unfolded protein response (UPR) is closely linked to autophagy, in that deficits in autophagy cause ER stress, and autophagy being an important pathway mobilised to clear unfolded and misfolded proteins by the UPR [266]. The UPR and has also been implicated in CD due to genetic variants in several key UPR genes being associated with CD (table 2.2). Cellular stress causes the accumulation of unfolded and misfolded proteins in the endoplasmic reticulum (ER), and is sensed by three transmembrane ER proteins: IRE1α, which acts via transcription factor XBP1; PERK, together with eIF2α, acting via transcription factor ATF4; and ATF6 which is cleaved freeing its cytosolic N-terminal region ATF6(N), which localises to the nucleus and functions as a transcription factor

[705]. These lead to broad downstream effects to counteract the ER stress and clear unfolded proteins, however if homeostasis is not restored then the UPR promotes apoptosis.

Gene	Function
ATG16L1	A core autophagy protein with important roles in Paneth cells [86], and one of the strongest genetic risk factors for CD [693, 239, 521, 193, 41, 643, 687, 554, 266, 162, 43], particularly ileal disease [654, 188].
IRGM	Involved in the induction of autophagy and clearance of intracellular bacteria including *Mycobacterium tuberculosis* [107] and is a well established risk gene for IBD [693, 193, 486, 41, 628, 687, 266, 162, 451, 532]
ITLN1	Like NOD2, ITLN1 responds to cell wall components (galactofuranosyl residues) of luminal microorganisms. Involved in autophagy, bacterial clearance and epithelial integrity [637], *ITLN1* is found in an IBD risk locus and contains an ileal CD associated SNP [41, 194, 297, 162]
LRRK2	Has GTPase and Kinase activity and is involved in autophagy [185, 678], NOD2 function in Paneth cells [733], NFAT activity [387], NFκB-dependant transcription, phagocytosis and bacterial killing [202]. The CD risk allele leads to reduced expression, and exacerbated DSS colitis in mice [387, 193, 239, 41, 628, 643, 687]. Also associated with Parkinson's disease, leprosy, and ankylosing spondylitis [630, 142].
MTMR3	Involved in autophagosome formation and phosphatidylinositol regulation in autophagy, and an IBD genetic risk locus [258, 266]
NOD2	Responds to MDP leading to autophagy and xenophagy via recruitment of ATG16L1 to bacterial entry sites in the plasma membrane and is the strongest individual genetic risk factor for CD [693, 239, 521, 193, 41, 687, 266, 43] (section 2.2.1)
ORMDL3	Regulator of the unfolded protein response and IBD risk locus [41, 422, 266, 284]

Table 2.2: Continued overleaf

Gene	Function
PTPN2	Autophagosome formation, regulation of autophagy gene expression (BECN1, ATG7, IRGM), response to MDP, TNFα and IFNγ, and clearance of intracellular bacteria, with function disordered by CD-associated genetic variants [553, 554] also seen in T1D and RA [693]. Well-established CD risk gene [693, 193, 486, 41, 643, 687, 553, 554, 43].
ULK1	Mammalian equivalent of yeast Atg1, ULK1, which contains a CD-associated SNP [256, 439], forms a complex with ATG13 and RB1CC1 which is central to the initiation of autophagy, particularly in response to starvation. Regulated by post-translational modification by MTOR, AMPK and TIP60 [698, 511].
Vimentin	A cytoskeletal intermediate filament protein which directly interacts with NOD2 and is required for localisation of NOD2 to the plasma membrane. CD-associated *NOD2* polymorphisms disrupt Vimentin-NOD2 binding [590, 459].
XBP1	Transcription factor activated by IRE1α as a core component of the unfolded protein response triggered by ER stress [310] *XBP1* is in a risk locus for IBD [240, 37, 665], its deletion causes spontaneous enteritis, and SNPs within *XBP1* are associated with CD and UC [310, 511].

Table 2.2: Autophagy and unfolded protein response genes which have been linked to Crohn's disease or IBD

2.3 Environmental risk factors

Indications of non-genetic risk factors Incidence and prevalence have continued to rise for several decades [436, 49] especially in children [259, 262] and purely colorectal disease has become more common [358].

The prevalence of IBD is highest in North America and northern and western Europe, but is becoming increasingly common in previously lower-incidence areas such as Asia, South America, southern and eastern Europe and north Africa [500, 455, 82, 83], and migration to a higher risk country increases risk of developing IBD [40]. Children of immigrants and non-immigrants have similar risk and a young age at migration increases the risk in first generation immigrants indicating an importance of early exposure [51]. Uncertainty regarding the true incidence of IBD in the originating country and differences between the migrant and native populations in seeking medical consultation complicate comparisons and there is also heterogeneity between migrant groups which may reflect differences in risk factors such as smoking [503].

Smoking Smoking is perhaps the most well-established risk factor for Crohn's disease, and yet areas of the world with the highest rates of smoking (up to 65% in adult African and Asian males) have low incidence of CD, and areas with the highest incidence have low smoking rates (27% in Canada and 19% in Sweden) [55]. Additionally smoking cessation is a risk factor for onset and exacerbation of ulcerative colitis [55] and smoking reduces the risk of developing PSC in both UC (OR 0.21) and CD (OR 0.17) [70].

Childhood immunity A common theme among many of the reported environmental risk factors for IBD is exposure of the immune system to various triggers during childhood. One aspect of this is the 'hygiene hypothesis', which was proposed in the late 1980s to explain the increasing prevalence of childhood asthma and eczema [591] beyond the rate explainable by genetic changes [489]. Increasing numbers of older siblings was found to protect against the development of various atopic conditions, and it was hypothesised that "allergic diseases were prevented by infection in early childhood, transmitted by unhygienic contact with older siblings" [591]. This has become an increasingly important theme in the modern understanding of allergy and has been proposed as a factor in autoimmune and inflammatory diseases [484] including IBD [334, 23, 206, 455]. Much of the interest now focuses on the effects of childhood environment on the intestinal

microbiota [489, 484].

Appendectomy has a widely reported strong protective effect against developing UC [39][341], and to predict a milder clinical course with a reduced risk of colectomy [129] but an increased risk of PSC [70]. Age <20 at appendectomy [196] and inflammation of the appendix rather than its removal [196, 46] appear to be particularly relevant. However, many potential confounding factors such as smoking and incipient IBD being misdiagnosed as appendicitis have been raised [195, 341, 577, 307] and other studies have failed to demonstrate the effect [577]. Studies in Crohn's disease demonstrate no effect [246, 577], a mild protective effect [84, 506], or an increased risk [307], leaving the importance unclear. Tonsillectomy has also been reported to reduce risk of both UC and CD [246, 84] but many studies have failed to show this effect [186, 431] or shown an increased risk [206, 706].

A meta-analyses of studies on breastfeeding showed a protective effect for both CD and UC overall [335] and in early onset disease [35], although some studies show no effect or an increased risk [39]. There is some evidence of a duration-response effect, complicating analysis [206].

Andrew Wakefield and colleagues proposed measles virus and measles vaccination as a risk factor for Crohn's disease [677, 676, 620]. Although this received less public attention than his claims for a link between MMR vaccination and autism, and the papers have not been retracted, they did receive considerable criticism [177, 523] and many studies found no link [495, 564] including a study he published [442].

Some studies have reported associations between IBD and vaccinations for Measles [246], Pertussis [246], Polio [246, 495] and TB [39]. Though in these studies vaccinations are highly correlated with year of birth and geographical origin, and there are problems with reporting, recall bias [84], and study heterogeneity [495].

The recommendation for patients with IBD is to receive normal vaccinations [106], however several studies have found low vaccination rates in IBD [724, 586]. This is concerning due to increased susceptibility to infection from immunosuppressive medication. The reasons for missing vaccinations in addition to those seen in the healthy population included contraindication due to concurrent immunosuppressive medication and disease flares at the time of scheduled immunisation, and gastroenterologists presuming that vaccinations would be managed in primary care [724, 586, 686].

Microbiota and infectious agents Recurrence of Crohn's disease following surgery is common [659], and depends on the presence of the faecal stream [537]. Presence of bacteria is necessary for inflammation in some animal disease models [561, 7, 150, 512, 417], and many IBD risk loci contain genes with roles in responding to bacteria (enrichment $p = 2.4 \times 10^{-20}$) such as *NOD2*, *IL10* and *CARD9* [297].

Mycobacterium avium paratuberculosis (MAP), which can be found in water supplies, meat, and pasteurised milk, causes Johne's disease in sheep and cattle; a fatal progressive chronic granulomatous enteritis with some similarities to Crohn's disease [388]. In his 1913 description of chronic interstitial enteritis, Dalziel noted the similarities between the two conditions

> In vol. xx of the *Journal of Comparative Pathology and Therapeutics*, McFadyen draws attention to Johne's disease, a chronic bacterial enteritis of cattle [...] In my cases the absence of the acid-fast bacillus would suggest a clear distinction, but the histological characters are so similar as to justify a proposition that the diseases may be the same [141].

In the 1980s unidentified mycobacteria were isolated from the intestines of CD patients and caused granulomatous disease when injected into mice and goats [114, 615]. Many studies have since demonstrated a significantly increased rate of MAP in CD patients [181, 1]. Even if MAP infection is just a consequence of the impaired barrier function and immune differences in CD with no role in pathogenesis, its presence shows a highly replicable correlation with CD [388, 181, 1, 29].

Another microorganism that has been linked to Crohn's disease is *Escherischia coli*, specifically adherent invasive *E. coli* (AIEC). CD patients have increased numbers of coliforms in faeces and associated with the intestinal mucosa, and the strains differ from those seen in acute gastrointestinal infections and the healthy bowel [143]. AIEC are strains characterised by their ability to adhere to intestinal epithelial cells and invade macrophages where they are able to survive and replicate [143]. AIEC are strongly associated with ileal CD, especially active disease, and are associated with relapse, greater need for surgery and early recurrence [412].

AIEC are present in some healthy individuals without causing disease implying that additional genetic or environmental factors found in CD may be necessary

to cause inflammation [412]. CD-associated variants in NOD2 reduce the ability to deal with AIEC by preventing interaction with vimentin and membrane localisation [590]. CEACAM6 and Gp96 both facilitate AIEC adhesion, and are overexpressed in ileal CD [146, 525, 412]. AIEC upregulate several microRNAs leading to a reduction in ATG5 and ATG16L1 and inhibiting autophagy [456], and Crohn's disease is linked to several other deficiencies in autophagy (section 2.2.2 & table 2.2).

Other organisms which have been linked to Crohn's disease include *Faecalibacterium prausnitzii* [584, 244], *Yersinia* spp., *Chlamydia* spp., *Listeria* spp., and *Bacteroides fragilis* [318, 388]. However, the ability to concurrently assay the proportions of many bacterial species and the lack of one clear microbial culprit for IBD has led to a focus on the overall microbiota rather than individual species.

Imbalance in the microbiota of the intestine, or dysbiosis, has been demonstrated in the gut [209, 190] and stool [480] in IBD, is detectable before treatment [209], worsens with more severe disease [480, 209], differs between CD and UC [480], and is not limited to areas of active inflammation [209].

The microbiota is acquired during delivery [281] and shortly after birth, maturing during the first several years of life [336], under influence from early childhood environmental factors [446] yet differing substantially in identical twins [640]. The healthy microbiota has numerous functions including immune homeostasis and development, GI development and metabolic functions [120, 190]. Factors which affect the established microbiota include smoking [485], alcohol [487], medications (particularly antibiotics), eukaryotic parasites, and diet [120]. The common findings in IBD are an overall reduction in diversity, with decreased Firmicutes and increased Proteobacteria [190, 480, 417, 244], though there is heterogeneity in specific findings.

Meta-analysis showed that antibiotics in IBD have a statistically significant benefit in inducing remission, preventing exacerbations and reducing fistula drainage [318], yet antibiotics have also been shown to increase the dysbiosis seen in IBD [480, 209]. Antibiotic exposure at a young age is associated with an increased risk of developing IBD [282, 263], though antibiotics may be given for incipient undiagnosed IBD, and the relative contributions of antibiotics, the infection that necessitated their use, or the immunological background which may have predisposed to infection are unknown.

Faecal transplantation has become established as an effective treatment for recurrent *Clostridium difficile* infection [656], and has garnered interest as a method to transform the dysbiosis seen in IBD to a healthy phenotype. While it is well

tolerated and there is some evidence of effectiveness, larger scale trials are needed [123, 283]. Likewise there is some evidence of benefit from the oral delivery of live organisms (probiotics), but the evidence is stronger for UC than CD and further research is required [541, 103]. The possibility of harnessing the immunomodulatory functions of helminths has drawn some interest, but also needs further research [691, 692, 595].

Other environmental risk factors A number of other environmental factors which modulate risk or severity of IBD have been reported, some of which such as westernization or latitude may be surrogates for other environmental risks or genetic factors. These, and the previously discussed risk factors are shown in table 2.3.

Factor	Effect	References
Smoking	↑CD ↓UC	[354, 88, 402]
OCP	↑CD ↑UC+smoking	[317, 126]
Breastfeeding	Protective	[500, 335, 35, 206, 39]
Appendectomy	↓UC	[159]
Helminths	Protective	[691, 692, 595]
Dietary fibre	↓CD	[14]
Weight	↑CD	[438, 315]
Exercise	↓CD	[316]
Sanitation	Deleterious	[159, 334]
Antibiotics	Deleterious	[500, 282, 263]
NSAIDs	Deleterious	[12]
Vitamin D	↓CD	[13]
Latitude	Deleterious	[23, 455]
Urbanisation	Deleterious	[169, 587]
Air Pollution	Deleterious	[15, 306, 45]

Table 2.3: Summary of evidence for exposures and environmental risk factors in CD and UC together (deleterious / protective), or individually (↑increased risk, ↓decreased risk). OCP: oral contraceptive pill, NSAID: non-steroidal anti-inflammatory drug

2.4 Epigenetic mechanisms

Epigenetics is the study of maintainable changes in gene function by mechanisms other than DNA mutation; "The structural adaptation of chromosomal regions so as to register, signal or perpetuate altered activity states" — in short "inheritance, but not as we know it." [66]. The most studied epigenetic mechanisms are DNA methylation, histone modification and microRNAs.

2.4.1 DNA methylation

The cytosine-guanine (CpG) dinucleotide occurs in vertebrate DNA at approximately 20% of the rate which would be expected in a random sequence of bases [65]. This may be due to the increased prevalence of 5-methylcytosine in CpGs (60–90% [545, 739]) compared to non-CpG cytosines and that 5-methylcytosine is prone to spontaneous deamination, leading to a point mutation to thymine (figure 2.1). CpG islands (CGIs), where CpG frequency can be 10-fold higher than the rest of the genome [65], and adjacent 'shores' [288] were hypothesised to have biological significance due their apparent preservation. CGIs mainly occur in 5′ promoter regions and are usually unmethylated [418, 44]; however, in cancer aberrant methylation and transcriptional silencing of tumour suppressor gene CGIs are frequently found [44].

DNA methylation causes downstream effects by interfering with the target recognition of DNA-binding molecules, or by recruiting methyl binding domain (MBP) proteins [79].

Tissue-specific methylation is more frequently found in intragenic CGIs (24.4%) than in 5′ promoter CGIs (2%) [418] and methylation within genes (intragenic methylation) may regulate alternative promoter regions [418] and influence splicing [569].

Artificially causing promoter CGIs to become methylated can cause transcriptional silencing and inactivated genes can be reactivated by inhibiting DNA methylation [65]. Alternatively, changes in gene regulation or expression, or underlying genetic differences may explain some of the described disease-associated methylation changes [558] making interpretation difficult and somewhat controversial.

A methyl group may be added to cytosine in CpGs by one of a family of DNA methyltransferase (DNMT) enzymes from the methyl donor S-adenosylmethionine (SAM) — requiring dietary micronutrients including folate, B12, and choline [18].

DNMT1 maintains the previous methylation state after DNA replication, whereas DNMT3a and DNMT3b produce new methylation. The process of DNA demethylation can occur passively by not maintaining hemimethylation during DNA replication, or actively by oxidation caused by ten-eleven translocation (TET) enzymes leading first to hydroxymethylcytosine (5hmC) and then to formylcytosine (5fC) and carboxylcytosine (5caC) which are excised by thymine DNA glycolase (TDG) and repaired by the base excision repair pathway (figure 2.1) [504, 48, 633].

Treating DNA with sodium bisulfite causes the conversion of cytosine to uracil (figure 2.2) unless there is a methyl or hydroxymethyl group. Hydroxymethylcytosine shows greater tissue specificity and is most prevalent in the central nervous system and embryonic stem cells [447, 346], but is less common and less well studied than methylcytosine. Hydroxymethylation is indistinguishable from methylation by standard techniques based on bisulfite conversion, however alternative techniques such as using TET and β-glucosyltransferase with bisulfite conversion can be used to detect both modifications. Hydroxymethylcytosine is more frequently found in regions undergoing active demethylation [588], potentially causing an underestimation of the true degree of hypomethylation reported in some findings.

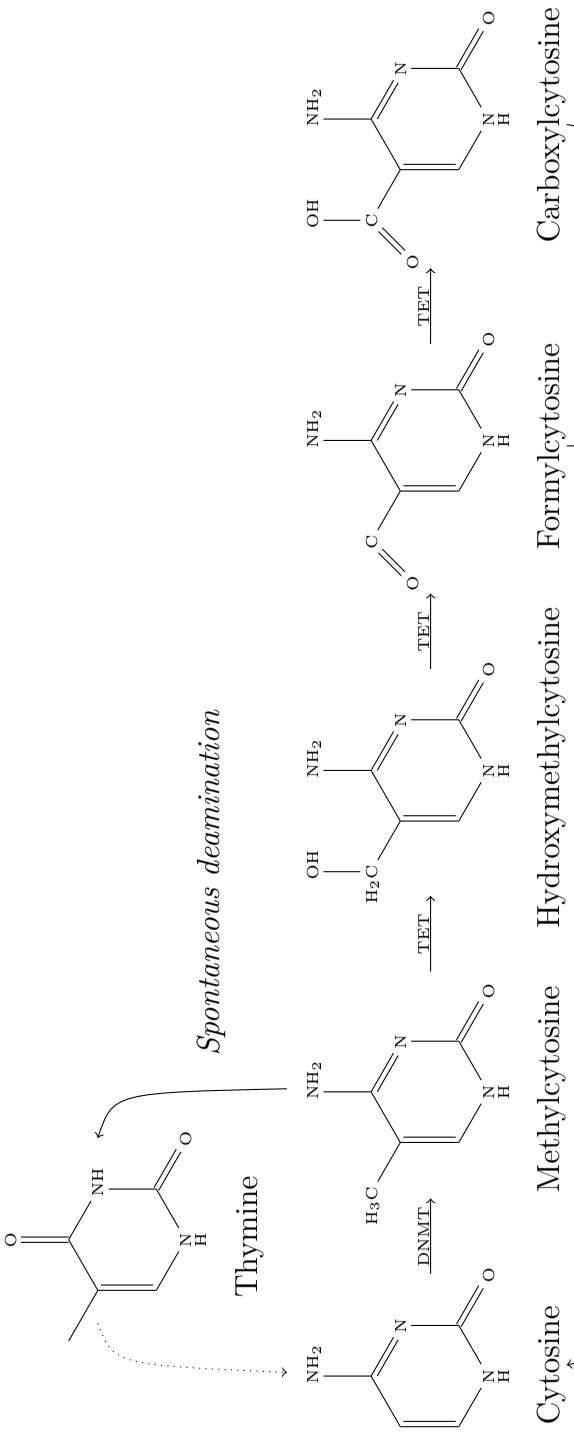


Figure 2.1: Conversion of cytosine to methylcytosine by DNMT, which can mutate to thymine by spontaneous deamination, and may be repaired by DNA repair enzymes. In active demethylation 5mC can be further oxidized by TET to form hydroxymethyl-, formyl-, and carboxylcytosine, which are excised and repaired.

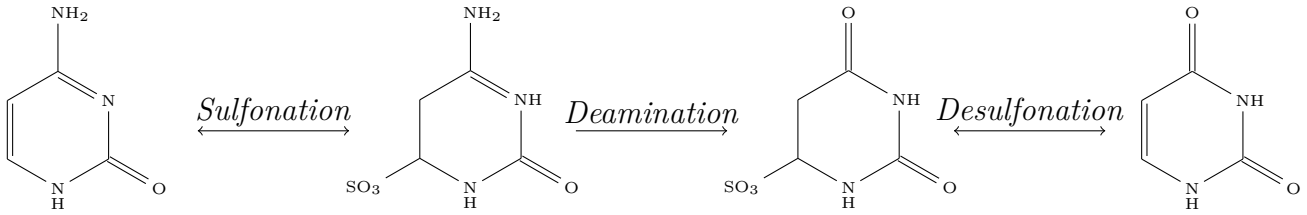


Figure 2.2: Conversion of cytosine residues to uracil by sodium bisulfite. Presence of a 5′ modification prevents this reaction under normal experimental conditions.

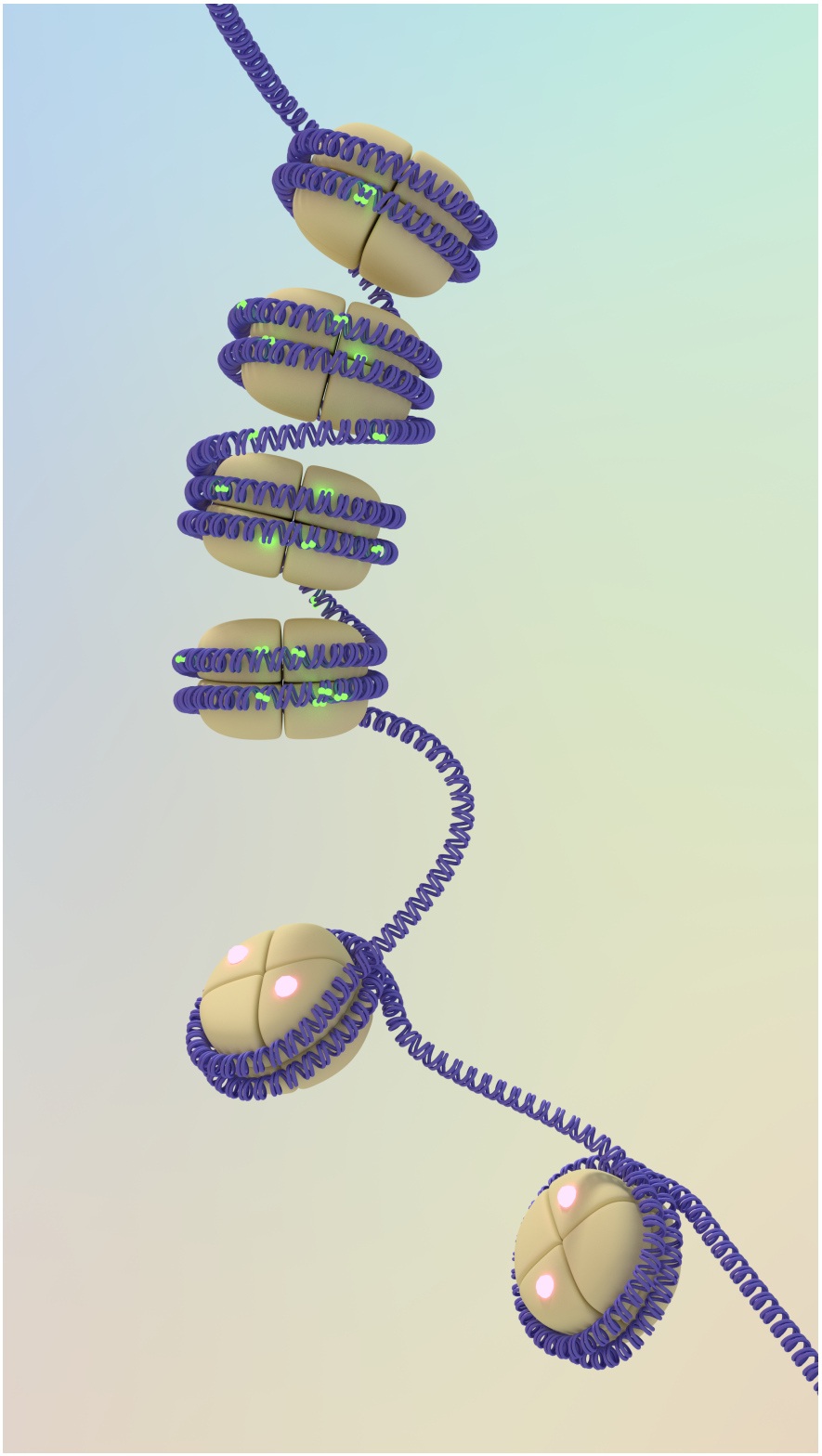


Figure 2.3: An illustration of DNA methylation (green) and histone modification (red) in heterochromatin (right) and euchromatin (left).

2.4.2 Histone modifications

The structural apparatus which packages the DNA double helix operates on multiple levels, the lowest of which is the wrapping of 145–147bp of DNA around nucleosome cores. Nucleosomes, comprising an octomer of histone proteins H2A, H2B, H3 and H4, occur approximately every 200bp in the eukaryotic genome and are the primary determinant of the accessibility of DNA to other molecules [397] (illustrated in figure 2.3). Each histone comprises a globular region and a protruding flexible N-terminal tail, which is subject to a number of post-translational modifications such as acetylation, methylation, ubiquitination, glycosylation and others [24], which are added and removed dynamically by opposing families of enzymes, making histone modifications more dynamic than DNA methylation.

These modifications alter chromatin structure, non-specifically influencing the accessibility of local DNA to DNA binding proteins and directly affecting the action of proteins which interact with histone tails themselves. Acting as an impediment to both DNA replication and gene transcription, the nucleosomes are disassembled in advance of these processes and replaced afterwards [81] with duplicated modifications [504]. In addition to direct accessibility of DNA, combinations of histone modifications are also correlated with changes in DNA methylation, and with differences in RNA expression separate from DNA accessibility and DNA methylation [350].

The presence and location of multiple modifications interact in a complex system, dubbed the 'histone code' [592]. The most studied individual modification in the context of IBD is histone acetylation; H3K27ac and H3K9ac, regulated by histone acetyltransferase (HAT) and histone deacetylase (HDAC) families of enzymes, and associated with promoter and enhancer activation, will be the only modification discussed at length in this review.

Humans have 18 HDACs, classified by their homology to yeast HDACs [709]. Class I (HDAC1, 2, 3 & 8) and II (HDAC4, 5, 6, 7, 9 & 10) are zinc based [709], class III (SIRT1–7) are dependant on NAD+ for lysine deacetylase activity [550] and do not share inhibitors with other HDACs, and class IV comprises just HDAC11, which has some similarities to classes I & II [201]. HDAC isoforms may vary by tissue, intracellular localisation and cell type [180, 709, 201], have a variety of non-histone substrates and non-deacetylase functions [709, 550] and HDAC inhibition only affects a small proportion of genes [409].

The endogenous short-chain fatty acid (SCFA) HDAC inhibitors are produced by microbial fermentation of dietary fibre [105]. Additionally compounds found in

garlic, broccoli, nuts, some leaves, dark cherries, red grapes and curcumin (found in turmeric) alter HDAC and HAT activity [715]. HDACs may also be affected by neuronal stimuli, immune activity, physical exercise, fasting, smoking and high salt diets [715] and are therapeutic targets for cancer.

2.4.3 MicroRNA

MicroRNAs are short, non-coding, single-stranded RNA molecules of approximately 22nt, which regulate expression of multiple mRNAs by guiding Argonaute proteins to partially complementary target sequences in mRNA [302]. There are 1881 annotated human microRNAs listed in the current version of mirbase[2] [344].

In the canonical pathway, microRNAs are processed from primary transcripts of several kilobases by the RNAse III enzyme Drosha to form a pre-microRNA hairpin of approximately 75nt, which is further processed by the RNAse III enzyme Dicer to produce the mature 22nt microRNA. The sequence complementary to the mature microRNA is usually degraded, but in some cases can also be a functional microRNA [42]. The evolutionarily conserved [374] 5′ seed region of the microRNA is almost always responsible for target recognition, binding to its target with greatly increased affinity due to the conformation induced by Argonaute interaction.

Alternative pathways such as the production of microRNAs from excised introns ('mirtrons') [352, 649], from the 3′ UTR of a host gene [517], small nucleolar RNAs [173, 428] and direct processing of microRNA precursors by Argonaute proteins [428] also exist. Additionally, microRNAs are not the only class of small RNAs to guide Argonaute proteins [327, 428] and small RNA targeted Argonaute mediated regulation does not target only mRNA [268, 428, 327].

MicroRNAs are regulated by endogenous RNAs (ceRNAs) [309] or 'sponges' — RNA transcripts with multiple target sites to reduce the availability of one or more microRNAs. MicroRNA sponging is a function for some non-coding RNAs [153], especially stable circular RNAs [546, 245]. Viral RNA has also demonstrated sponge activity [378, 407] to target host antiviral microRNAs.

Expression and activity of some microRNAs is modified by other epigenetic mechanisms [543, 241], genetic variants [74, 117], immune activation [736], inflammation [704], microorganisms [140, 456], viruses [378, 407], and parasites [78].

Approximately half of MicroRNAs are found in known genes (normally in the

[2] `www.mirbase.org` Version 21, published June 2014

sense strand) [215] and may be co-transcribed or have independent promoters and 20% of these were found to either directly or functionally antagonise their host gene [265].

MicroRNAs have been shown to have roles in development, T cell differentiation and signalling, immune mediation, intestinal homeostasis, autophagy and inflammation [536, 302] and have been implicated in autoimmune diseases [492] and cancers.

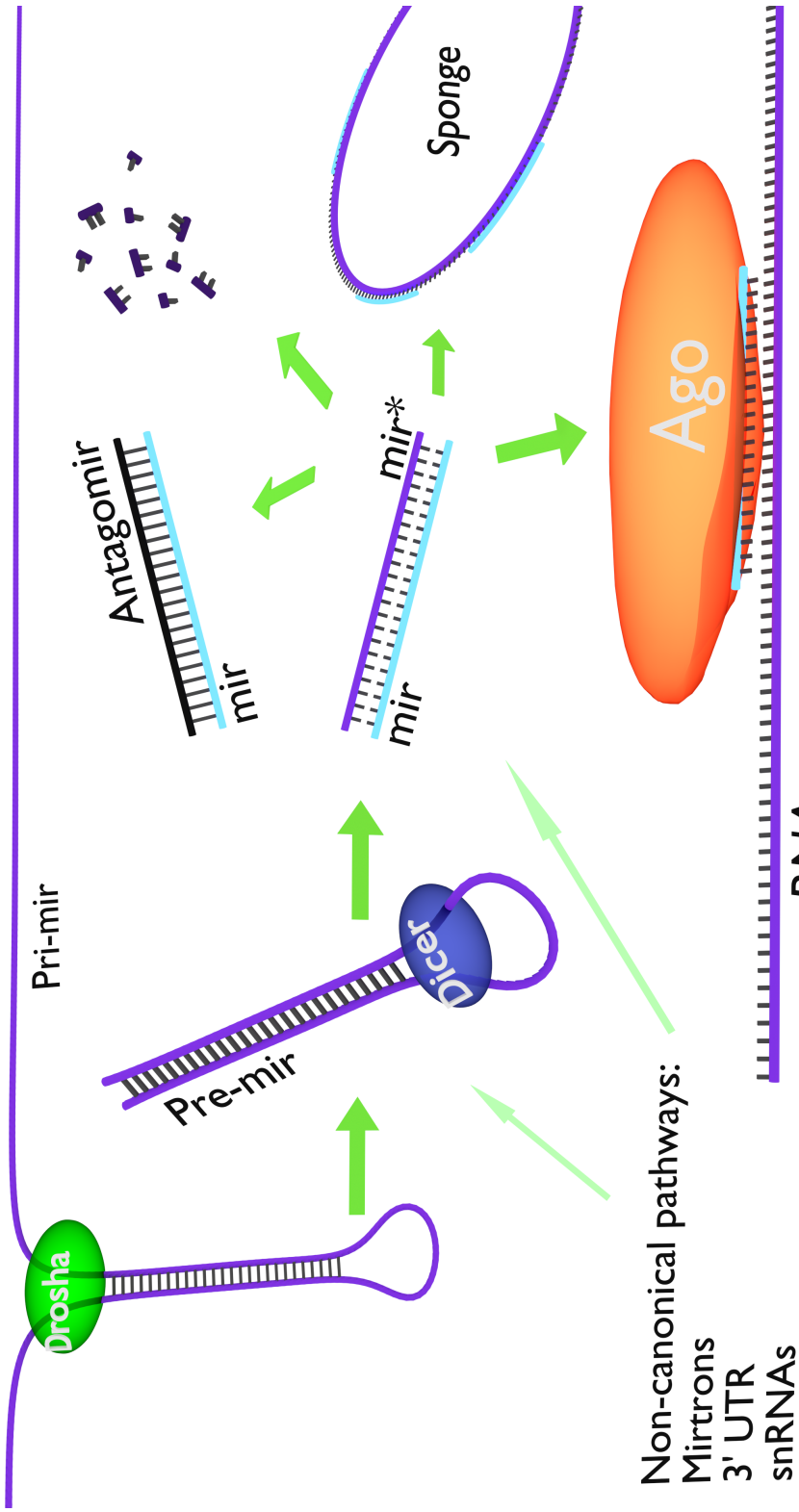


Figure 2.4: MicroRNA action and processing pathways

2.5 Epigenetics in Crohn's disease

2.5.1 Findings in IBD

Targeted DNA methylation studies Early studies examined overall levels of DNA methylation, finding global hypomethylation in rectal mucosa of patients with long-standing UC [214]. Following work looked at small numbers of selected genes, with a focus on UC and UC-associated colorectal cancer (UC-CRC). Small study size and the inherent heterogeneity of cancers have led to some conflicting results, but also a number of reproducible findings.

Methylation changes affecting expression of the anti-tumour *CDKN2A* products p16ink4a and p14ARF were among the first reported findings. p16INK4a inactivation by promoter hypermethylation was seen in sporadic CRC and was strongly associated with long-standing and extensive UC [271]. These regions were repeatedly found to be hypermethylated in UC [441], and increasingly methylated in non-dysplastic and dysplastic tissue of patients with UC-associated dysplasia [289]. p14ARF methylation was shown to be an early event in UC-CRC [547].

E-Cadherin (*CDH1*) promoter methylation, associated with mucosal inflammation in UC [542], was found in spontaneous and UC-associated CRC [694] and was hypothesised to have a role in the progression of long-standing UC to neoplasia [31].

ER (oestrogen receptor) gene methylation is increased in UC and further increased in UC associated cancer, additionally methylation was greater in non-neoplastic tissue of UC patients with cancer compared to UC patients without cancer [289, 627, 198, 22] and has been proposed as a biomarker for risk of cancer in UC [627].

CEACAM6 is abnormally expressed in the colonic mucosa of CD patients. CEACAM6 expression is regulated by methylation in mice, and a methyl donor deficient diet leads to increased expression with resultant susceptibility to colonisation by AIEC and inflammation [146].

A number of other genes have also been shown to be epigenetically dysregulated in UC (*CXCL14*, *CXCL5*, *GATA3*, *IL17C*, *IL4R* [308] & *TUSC3* [22]), or between UC patients with and without dysplasia (*MYOD1*, *CSPG2* [289], *RUNX3*, *MINT1*, *COX-2* [203], *SLIT2* [32], and *HPP1* [548]).

Methylation of *CDH1*, *MYOD1*, *GDNF*, *HPP1* [542] & *SLIT2* [390] correlated with disease activity, and *MDR1* [596] & *PAR2* [597] with disease phenotype.

IFNG SNP rs1861494 (C/T) disrupts a CpG site. For the C allele, the CpG is 60% methylated and the adjacent CpG (57bp distant) was also hypermethylated compared to T allele carriers. Methylation of these CpGs was highly correlated with *IFNG* promoter methylation, and nucleoprotein binding. The T allele is associated with a more severe phenotype in both UC and CD, and with higher IFNG protein expression [218]. It had previously been shown that *IFNG* methylation differs in IBD [219] and that *IFNG* hypomethylation correlates with IFNG secretion, risk of surgery, and increased antibody response to microbial antigens [217].

Polymorphisms in *MDR1* [596], *XRCC1*, *GSTP1*, *GSTT1* [598], *CD14*, *IL-1β*, *p22PHOX*, & *MBL2* [599] have also been found to correlate highly with certain methylation changes in UC.

Epigenome-wide association studies A number of epigenome-wide association studies (EWAS) have been published in the past five years (table 2.4) coinciding with advances in technology allowing characterisation of methylation changes at high (single-nucleotide) resolution at an increasing number of sites (10^3–10^5). The published EWAS studies are not amenable to meta-analysis, as they are mostly small studies, with varying tissue types, analysed on a variety of platforms, by differing techniques. EWAS are subject to complications not seen in GWAS due to epigenetic plasticity (in response to inflammation, age, smoking, drugs etc) and the fact that epigenetic markers can vary extensively between cell types [350].

Epigenetics and GWAS risk loci CD-associated methylation changes are enriched in proximity to GWAS risk loci [458, 4], as are T cell subpopulation specific transcriptional differences [508]. Allele-specific methylation has been demonstrated to be genetically regulated by variants associated with complex diseases including IBD and may provide further insight into the observed link between DNA methylation and GWAS risk loci [280].

H3K27 acetylation, a marker of active regulatory elements, is significantly enriched in proximity to GWAS risk loci for IBD (2.5 – 3.5 times enrichment, 92/163 risk loci) [434]. H3K27 profiles differed between intestine and immune cells, with immune cell epigenetic regulatory elements colocalising with risk loci for CD or UC, and intestinal cell regulatory elements enriched in proximity to UC risk loci and diminished in proximity to CD risk loci [434].

Study	Platform[a]	Tissue	Samples[b]	Results[c]	Significance[d]
Nimmo 2012 [458]	27k	Blood	21CD 19HC	1117	FDR<0.01
Harris 2012 [250]	450k	Blood	17CD, 11UC, 20HC	1	FDR<0.2
Lin 2012 [381]	GG	B cells	9CD 9UC 18HC	11	uncorrected
Adams 2014 [4]	450k	Blood	36CD 36HC	65 CpGs, 19 DMRs	Bonferroni, FDR
Lin 2011 [380]	GG	Intestine	9CD 17UC	7IBD, 25CD, 13UC	uncorrected
Häsler 2012 [251]	27k	Colon	10 UC/HC twins	61 Genes	FDR
Cooke 2012 [125]	27k	Rectum	16CD 16UC 8HC	3604UC 472CD	FDR
Koizumi 2012 [337]	MS-AFLP	Colon	14UC 11HC	251 Probes	Holm
Harris 2014 [249]	450k	Colon	15CD 9UC 22HC	182CD 3365UC DMRs	$p<10^{-4}$
Kraiczy 2015 [345]	450k	IEC	10CD 10UC 10HC	233	Šidák<0.01
McDermott 2015 [420]	450k	PBMCs	88CD 61UC 39HC	3196CD 1481UC	Bonferroni

Table 2.4: Published IBD EWAS studies. [a]27k: Illumina HumanMethylation27 BeadChip, 450k: Illumina HumanMethylation450 BeadChip, GG: Illumina GoldenGate Genotyping Assay, MS-AFLP: methylation sensitive amplified fragment length polymorphism. [b]HC: Control [c]CpGs unless otherwise specified [d]<0.05 unless otherwise specified, FDR: Benjamini-Hochberg false discovery rate

Histone modifications There are fewer data concerning alterations to histone modifications in IBD than DNA methylation changes, possibly due to increased technical difficulty and reduced scalability of experiments, the wide range of histone modifications and increased plasticity compared to DNA methylation.

Dynamic changes in H3K9 methylation have been shown to regulate transcription of certain inflammatory genes [540], and there is evidence for regulation of NOD2 [369] and TLR4 [601] expression by histone acetylation. Histone acetylation [635] and methylation [604] are significantly disrupted by experimental colitis models and HDAC expression has been shown to respond to inflammation in UC [461] and exposure to certain microbes [720].

Deletion of both *HDAC1* and *HDAC2* in murine intestinal epithelial cells (IEC) led to chronic inflammation [638] and increased sensitivity to DSS colitis [639], whereas mice with HDAC2 deficient IEC did not have chronic inflammation and were protected against DSS colitis [639].

HDAC3 expression is significantly decreased in IEC from IBD patients compared to healthy controls. Deletion of *HDAC3* in IEC of mice caused alterations in gene expression and microbiota, loss of Paneth cells, impaired epithelial barrier function, inflammation and increased response to DSS. These changes were dependant on the presence of commensal bacteria [7].

Mice with intestinal *SIRT1* knockout show increased resistance to experimental colitis and colitis-induced neoplasia [389].

Histone acetylation, by inhibition of HDAC enzymes has a general anti-inflammatory effect, including protecting against and promoting resolution of experimental colitis models [211], hence making HDAC inhibition an attractive therapeutic target [180].

MicroRNAs A large number of microRNAs have been reported as either up or downregulated in IBD (see Kalla *et al.* [302] for a recent review) in response to immune stimulation [736], inflammation [704], microbial [456] or genetic factors [74, 117]. MicroRNAs are involved in regulation of both the innate and adaptive immune system, as well as numerous other physiological and pathological processes. Of particular interest in IBD are their roles in regulating NOD2 and toll-like receptors, T cell differentiation, autophagy, cytokine release and inflammation [302].

IBD related products and pathways linked to microRNAs include ATG16L1 (mir30c, 130a [456], 142-3p [730], 106b and 93[393]), NFκB (mir21 [736]), STAT3 (mir124 [340] and 21 [552]) and NOD2.

NOD2 activation induces microRNAs 29a, 29b and 29c, leading to a down-regulation in IL-23 and T_H17 response. CD-associated *NOD2* polymorphisms failed to induce mir29 and DSS colitis was worsened in mir29 deficient mice [74]. NOD2 also induces mir155 [74] and mir146a [210]. *NOD2* is itself subject to regulation by microRNAs: mir122 [111] 192, 495, 512 & 671 [117] directly inhibit *NOD2* by binding to a sites in the 3′ UTR, where disease-associated polymorphisms may interfere with microRNA binding [117].

Another notable microRNA known to be dysregulated in IBD is mir21, which is known to be overexpressed [565, 704, 600, 703, 725, 4] and hypomethylated [4] in IBD and *MIR21* knockout mice have been shown to be protected from DSS-induced colitis [565]. mir21 was one of the first described, and most frequently dysregulated microRNAs and is regulated by STAT3, BCL6 [552] and NFκB [736]. It is involved in T helper cell differentiation [552] with activity in other immune mediated diseases [492] and cancers [347] where it has been proposed as a prognostic biomarker [683] including predicting poor prognosis in colorectal cancer [457].

2.5.2 The potential of epigenetics as a diagnostic tool in IBD

DNA methylation biomarkers Aberrant DNA methylation as a diagnostic biomarker for cancer is being pursued, for example *SEPT9* in colorectal cancer [227, 685, 119] and *GSTP1* in prostate cancer [655]. Despite subtler methylation changes in complex diseases, multiple DNA methylation changes have been proposed as biomarkers.

A study of paediatric CD[3] (n=71) found multiple diagnostic biomarkers based on leucocyte DNA methylation at 2 loci capable of distinguishing between disease and symptomatic control with high accuracy (AUC<0.98) [4]. A combined assay of *BMP3* and *NDRG4* DNA methylation in stool detected 100% of cancer and high-grade dysplasia, and 67% of low-grade dysplasia at 89% specificity [332] in patients with inflammatory bowel disease (n=54). Other studies have demonstrated the feasibility of detecting DNA methylation levels in serum [34], intestinal biopsies [326] and stool [32] and have shown the possibility of biomarkers for disease activity [542, 390] and phenotype [596, 597].

MicroRNA expression biomarkers A large number of microRNAs have been proposed as biomarkers for a range of conditions, including neoplastic [697], ischaemic, infective, inflammatory, degenerative, psychiatric, and traumatic [234].

Pritchard *et al.* found that for 79 microRNAs reported as biomarkers for solid tumours, 58% were highly expressed in blood cells, and 91% of these were expressed at high levels in healthy plasma. For these microRNAs corrections for the variations in cell populations are necessary to ensure that the reported microRNA is not a proxy measurement for a shift in cell populations, altered rate of haemolysis, or secondary to malignancy or its treatment [502].

Haider *et al.* examined 192 microRNAs reported as biomarkers for 57 non-neoplastic conditions, including IBD, and found that small study size, variable normalisation and analytical procedures, and failure to replicate findings were common in the literature. They also point out that many reported microRNA biomarkers are expressed in numerous cell types, lack a strong rationale for the cell types assayed given the focus of the disease, and lack specificity. 24% were also reported as neoplastic biomarkers and 6 microRNAs were reported as biomarkers for 9 or more diseases (mir21, mir16, mir146-a, mir155, mir126 and mir223) [234], all of which have been reported as significantly dysregulated in IBD, if not

[3]Presented in chapter 6

explicitly as biomarkers.

Despite the general lack of specificity to a particular disease, some microRNAs have been reported to demonstrate differences in expression between CD and UC, and to vary by disease site [703, 702]. At this stage, microRNAs have been identified in blood [481, 702], serum [725] and biopsy tissue [704, 703, 128] which are differently expressed in IBD, but none have been developed to the point of clinical utility as biomarkers.

Detecting IBD-associated cancer A field defect, or field for cancerization is where cell division creates a local tissue environment sharing mutations, with subfields containing later acquired mutations, eventually producing cancerous lesions with accumulation of sufficient mutations. This explains the development of multiple primary tumours, and presence of pre-cancerous lesions near tumours. Fields where the defect is epigenetic rather than genetic have been demonstrated in Barrett's oesophagus [164], gastric epithelium [460, 278, 338] and bowel [289, 337, 271]; however, the shared epigenetic dysregulation is primarily due to shared local exposures rather than inheritance. Epigenetic field defects have been shown to develop in response to DSS exposure [312], microorganisms such as *H. pylori* [644, 460, 278, 113] and Epstein-Barr virus [594] and chronic inflammation including IBD [433, 289].

The epigenetic changes in field defects include specific methylation changes of tumour suppressor genes or oncogenes [165], histone modifications [604], or more general effects such as global hypomethylation [165, 214] or disruption of chromatin remodellers [605].

The presence of epigenetic abnormalities in the development of cancer has been well established, and is detectable prior to histological change [433]. As patients with IBD who develop cancer have detectable epigenetic differences present in both neoplastic and non-neoplastic tissue epigenetic biomarkers have been proposed to detect current or imminent cancer [332, 99, 627, 26].

2.5.3 New potential treatments from epigenetics

DNA methylation Nucleoside analogues are the most developed approach for DNMT inhibition. These drugs are incorporated into DNA in place of cytosine and due to the presence of a nitrogen at position 5 irreversibly bind to DNMT leading to its degradation [187]. Examples include 5-Azacytidine (AZA, Azacitidine) and 5-Aza-2′-deoxycytidine (DAC, Decitabine), which are used in the treatment of haematological conditions including myelodysplastic syndrome (MDS), acute myeloid leukaemia (AML) and chronic myelomonocytic leukaemia (CMML). There is evidence of synergistic activity when combined with HDACi such as TSA [92] and valproate [187]. Similar compounds designed to improve speed of onset [269], stability and delivery [187] are also being investigated.

Reversible DNMT inhibitors which do not require incorporation into DNA are being sought, and anti-DNMT actions of several existing drugs (hydralazine, procainamide, procaine, disulfram [379] and antidepressants [740]) natural compounds (EGCG, genistein, laccaic acid) and novel compounds (RG108, SGI-1027) [400, 187, 226] have been tested with varying results.

DNMT expression is known to be regulated by microRNAs, and both synthetic oligonucleotides (MG98) [187] and mir29 [174, 187] are a potential alternate mechanism for modulating DNMT activity, additionally there are numerous beneficial effects of mir29 unrelated to DNMT [204, 74].

DNMT3A [194] and *DNMT3B* [297] genes have been identified as IBD risk loci by GWAS, DNMT expression is increased in active UC [542], and many DNA methylation changes have been found in IBD; however, there are obstacles to therapeutic targeting of DNA methylation directly in IBD. DNMT inhibitors have direct cytotoxic effects [551, 473, 383] and so their activity in malignancy is not entirely epigenetic and there is insufficient data to draw any conclusions about long-term risks of DNMT treatment, including the risk of prompting cancer development by causing global hypomethylation [165]. Tumour suppressor gene CpG island methylation is more susceptible to treatment with AZA and DAC compared to gene body CpG islands [699], however DNMT inhibition is largely untargeted, and methylation changes in IBD show both hypo- and hypermethylation making DNMT inhibition of unclear potential benefit in IBD. Targeting DNMT inhibition to a particular cellular population, or targeted DNA methylation of specific pro-inflammatory or oncogenic genes using a DNMT enzyme linked to a DNA binding domain or zinc finger protein [301] may provide a more controllable therapeutic approach with further development.

Folate DNA methylation is dependant on the methyl donor SAM, the production of which requires micronutrients including B12 and folate, B2, B6, betaine, choline and methionine [18]. B12 and folate deficiencies are rare in newly diagnosed children in the USA [8, 261, 571] but are common in established IBD patients [228, 670], as are B2 and B6 deficiency [660]. There are multiple possible causes for these deficiencies, including anorexia, malabsorption, haemolysis, bacterial overgrowth, surgical resection and increased epithelial turnover [660, 228].

Rats fed a diet without B12, folate and choline had a significantly increased response to DSS exposure [108], and mice treated with folate supplementation were protected against *Helicobacter*-associated global hypomethylation, inflammation and gastric cancer [216].

Treatment of anaemia with B12 and folate supplementation is common in IBD [670] and several small studies [559] have shown non-significant trends towards a protective effect of folate against CRC and a dose-response effect. Further research into the effects of folate and other micronutrients on DNA methylation and cancer risk are needed, particularly as over-supplementation may increase cancer risk [328, 651, 21].

HDAC inhibitors Histone acetyltransferases, and enzymes which add or remove other histone modifications could also prove valuable targets, however research to date has heavily focused on HDAC inhibitors. There are several classes of antagonists for class I & II HDACs: hydroxamates (Vorinostat (SAHA), trichostatin A (TSA) and Givinostat), aliphatic acids (valproate, butyrate), cyclic peptides (Romidepsin) and benzamides (Entinostat) [400]. This review will focus on inhibitors of non-sirtuin HDACs, however, multiple classes of sirtuin inhibitor exist and are being pursued for a range of conditions [362]. Sirtuins are also inhibited by nicotinamide, a byproduct of their histone deacetylase activity [30] and nicotinamide has been shown to ameliorate bacterial and DSS induced colitis in mice [59] improve neutrophil killing of *S. aureus* by up to 1000-fold [351] and is used as a treatment for acne and dermatitis. Additionally, some HDACs are predicted to be regulated by microRNAs [180], and could be a target for microRNA focused therapeutics.

This is a large variety in the targets of HDAC enzymes and their inhibitors. However HDACi share anti-inflammatory effects, due to reduced inflammatory cytokine production, and regulation of NFκB and other transcription factors [211, 212, 180] and T cell differentiation [180]. HDACi are also pro-apoptotic and anti-cancer [211, 212, 180] and are involved in intestinal architecture, epithelial barrier

function, immune cell infiltration, expression of various inflammatory genes and microRNAs [638, 180]. Common side effects include fatigue, nausea, vomiting, diarrhoea and thrombocytopaenia [189, 180], less frequently QT prolongation or arrhythmia were reported, but resolved spontaneously with dose reduction or treatment cessation.

Oral administration of HDAC inhibitors valproate, SAHA or Givinostat protected mice from DSS-colitis and TNBS-colitis by reducing expression of pro-inflammatory cytokines including TNFα and IFNγ, inhibiting NFκB, and at higher doses inducing apoptosis [211, 212]. Anti-inflammatory effects of SAHA were observed at lower doses than required for anti-tumour activity [368].

A small open-label trial of oral Givinostat 1.5mg/kg/day in systemic-onset juvenile idiopathic arthritis reported an improvement in signs and symptoms, but no change in ESR and CRP and acceptable safety profile and tolerability [671]. A single RCT of Givinostat was undertaken in moderate-to-severe active CD but was terminated for futility after interim analysis of 40/80 patients[4].

Butyrate Factors produced by epithelial cells (e.g. TGFβ, IL-10, Retinoic acid, and prostaglandin E2) promote tolerance to microbial triggers [555], and the healthy microbiome produces the short chain fatty acids butyrate, propionate, and acetate [105] by anaerobic carbohydrate metabolism.

Butyrate is implicated in mucin release, electrolyte, water and pH regulation, acts as an energy source for colonocytes, promotes epithelial tight junction integrity and regulates cellular proliferation and differentiation [496] including driving regulatory T cell differentiation [200].

Butyrate demonstrates the general anti-inflammatory effects seen with other HDACi [496, 105, 398] and other anti-inflammatory effects including abolishing LPS-induced TNF secretion by PBMCs due to inhibition of NFκB expression and stabilisation of IκBα [560], GPR109A signalling (also caused by niacin) leading to IL-10 and IL-18 production and Treg differentiation [575], and reducing ROS production by neutrophils [385].

Butyrate producing bacteria are depleted in IBD [190] and colorectal cancer [681]. Butyrate oxidation and transport are disrupted in IBD [618, 145, 144] with evidence of normalisation following anti-inflammatory therapy, indicating the changes are caused by inflammation [145] and there is impaired sensitivity to the anti-inflammatory effects of butyrate in IBD [343].

[4]https://clinicaltrials.gov/show/NCT00792740

There is some positive evidence for oral butyrate [667], butyrate enemas [236, 237], fibre supplementation [688], and propionyl-L-carnitine to facilitate SCFA transport [429]. However, study size and design have been criticised [688] and effectiveness remains questionable [361].

MicroRNA-based therapies Potential microRNA therapeutics could reduce the expression or activity of a microRNA, increase expression of an endogenous microRNA, or supply non-endogenous short non-coding RNAs. There are difficulties inherent in microRNA-based therapeutics: targeting, and reaching the site of action, resisting degradation, generating an immune response, side-effects due to the numerous targets of microRNAs, and risk of cancer from gene therapy [205, 610, 33]. However, the pharmacokinetics of oligonucleotide drugs [622], and chemical and sequence modifications to increase half-life and reduce immunogenicity [470] are generally independent of the specific sequence, making this a potentially highly adaptable drug class.

Antagomirs are antisense oligonucleotides, modified to increase stability, which bind to target microRNAs causing degradation or sequestration. In the first phase 2 study of an antagomir in humans, antimir122 (Miravirsen) was given to 36 patients with chronic HCV infection in 5 weekly injections. The treatment was well-tolerated, with no discontinuations or dose-limiting effects [293]. The drug is a complementary sequence to mature mir122, with nucleotides modified to enhance binding affinity and increase resistance to nuclease degradation. Antimirs to microRNAs implicated in IBD such as mir21 [585] are in an earlier stage of development, but a potential use in various cancers may add impetus.

Similarly, oligonucleotides can be delivered to mimic an endogenous microRNA [33, 260] or with a sequence designed to target a specific mRNA. Delayed release preparations combined with permeability enhancers (e.g. sodium caprate) even allow oral delivery [622]. A phase 2 trial of an oral *SMAD7* antisense 21-base phosphothioate oligonucleotide (Mongersen) in Crohn's disease [437] showed a statistically significant induction of remission with good side-effect profile.

A further application of delivering short, stabilised oligonucleotides is to alter the proportion of transcripts. Splice-switching oligonucleotides (SSO) can be designed to promote skipping exons during translation [568] which for example allows a reduction in the membrane-bound functional TNFR2 and an increase in a secreted Δ7TNFR2 form which antagonised TNFα highly effectively in mouse disease models [222].

MicroRNA sponges could be used to reduce levels of target microRNAs, with

the added advantages of stability and ability to target multiple microRNAs. Viral and non-viral delivery mechanisms are being investigated, with the majority of animal experiments to date using viral vectors [610].

Some research has found dietarily ingested plant microRNAs are absorbed and active [732], though an attempt to replicate these findings failed, and the premise is controversial [110, 151]. Alternatively, immunomodulatory microRNAs or sponges produced by bacteria [140], viruses [378, 407] or parasites [78] could be introduced to the gut by selected or engineered organisms.

2.6 Paediatric IBD

The study presented in the following chapter was undertaken in children, concordantly it is important to consider the differences between this group and adult-onset IBD.

Up to 25% of IBD cases present in childhood or adolescence [370]. The incidence of paediatric IBD is rising and the average age of onset decreasing [259, 50]. The Scottish population has seen a dramatic increase in IBD of 76% from 1990–1995 to 2003–2008, with a significant reduction in the age of onset (13.2–12.1 years, $p<0.001$) of Crohn's disease [259]. The trend of increasing incidence in Scottish children can be seen since 1969, though methodological differences and changing diagnostic techniques make robust comparisons with early studies difficult [259].

The Montreal classification for Crohn's disease includes three factors, age at diagnosis, location, and behaviour, with age of diagnosis separated into $\leqslant$16, 17–40 and >40 [573]. However, there are significant differences between very early onset (VEO, age <10) and early onset (EO, age 10–17) IBD [469].

Most IBD GWAS studies were undertaken in adults, though some studies have looked at paediatric disease or compared paediatric with adult-onset disease [349, 284, 124]. There is a large overlap between the loci found in paediatric and adult diseases [258, 469], suggesting that genetic differences responsible for paediatric disease specifically may be poorly detected by GWAS, i.e. rare variants with large effect sizes, a result of complex epistatic interactions, or that non-genetic factors may be responsible for earlier onset. Some rare mutations in genes for IL-10, and IL-10 receptors have been found to be particularly associated with paediatric disease [213, 567, 57, 339]. In cases where the strength of a disease association varies strongly between children and adults, and the risk variant is also correlated with a particular disease distribution (such as *ATG16L1* with ileal disease) the order of causality may be difficult to discern [654].

In younger children ileal disease is much rarer and strongly associated with *NOD2* mutations [425, 371], family history for IBD and autoimmune conditions are more common [531, 498], and there are differences in serological markers such as ASCA[5] and anti-CBir1[6] [408]. Paediatric disease affects a higher proportion of males, the course is more severe and progressive with an increased requirement for immunosuppressants, surgery, or total parenteral nutrition, reduced response

[5]Anti–*Saccharomyces cerevisiae* antibody
[6]Anti–flagellin CBir1 antibody

to immunosuppressants and increased risk of death [531, 93, 494, 498, 223, 653]. However, surrogate markers of severity such as surgical and medical treatments can be difficult to interpret due to differences in physician and patient preferences between children and adults [653].

Children with IBD (particularly CD) are at risk of delayed growth and puberty due to malnutrition, anorexia, steroid treatment and the effects of inflammatory cytokines on growth hormone and insulin-like growth factor [224]. Impaired growth is present in more than 35% of paediatric CD patients, and failure to thrive was present at diagnosis for 44% of CD patients under 5 years old [324].

The Paris classification therefore modified the Montreal classification to include age at diagnosis <10 and growth failure, as well as several additions to classification of location and behaviour [370] (see table 2.5)

In addition to these factors, young people with Crohn's disease may miss schooling and exams, though without affecting measured outcomes or entry to tertiary education [419, 183, 401], are at an increased risk of depression, anxiety, body image issues, sleep disturbance, and negative impacts on social and family functioning [401, 324]. Many of these outcomes have been shown to be worse in Crohn's disease than other chronic illnesses [401, 324].

There are advantages and disadvantages to using children as a study population. With a more severe phenotype and a higher genetic component, genetic variants could be easier to detect, but paediatric disease could also be due to individually rare but highly-impactful mutations which are difficult to discover with GWAS. With regard to epigenetics, children have a far smaller exposure to factors which affect epigenetic markers such as smoking, comorbidities, polypharmacy, and age, which could make it easier to find a signal amongst noise. Conversely, if they have a higher genetic load than adult patients, then the non-genetic factors such as epigenetics could be of less importance in this group.

Montreal		Paris	
Age at diagnosis (A)			
		A1a:	<10
A1:	⩽ 16	A1b:	10–16
A2:	17–40	A2:	17–40
A3:	>40	A3:	>40
Location (L)			
L1:	Terminal ileum	L1:	Distal 1/3 Ileum
L2:	Colonic	L2:	Colonic
L3:	Ileocolonic	L3:	Ileocolonic
L4:	Upper GI	L4a:	Proximal to LoT
		L4b:	LoT – distal 1/3 ileum
Behaviour (B)			
B1:	Nonstricturing & nonpenetrating	B1:	Nonstricturing & nonpenetrating
B2:	Stricturing	B2:	Stricturing
B3:	Penetrating	B3:	Penetrating
		B2B3:	Stricturing & penetrating
Growth (G)			
		G_0:	No delay
		G_1:	Delayed

Table 2.5: The Montreal [573] and Paris [370] classifications of Crohn's disease. +L4 and +p (perianal) can be added as modifiers to any location or behaviour respectively. LoT: Ligament of Treitz

2.7 Conclusions

Despite extensive study of the genetic risk factors for IBD, a large portion of the disease variability remains unexplained. Epigenetics was predicted to offer some insight into the 'missing heritability' and the interplay between factors such as environmental exposure, gut microflora, diet, smoking and background genetic risk. Early studies have identified many epigenetic disturbances, and indeed abnormalities with core epigenetic mechanisms associated with IBD. In addition, understanding of the importance of epigenetic mechanisms to biological pathways critical to IBD is increasing.

Epigenetics remains a developing field, but increasing interest and technological developments should improve characterisation of epigenetic changes in IBD and increase understanding of their causes and interrelatedness. In the immediate term, these findings can be expected to provide insight to IBD researchers outwith epigenetics.

In principle, epigenetic marks have strong potential as biomarkers in IBD and other complex diseases as they reflect a range of biologically significant influences (such as disease activity, genetics, environmental, microbial and dietary factors). Unlike genetic risk factors they are changeable, but may be more stable than measuring expression. A lack of disease-specificity has been noted [502, 234], however when considering the nature of epigenetic regulation, with large numbers of targets, complex interactions and multiple feedback mechanisms this is not surprising. Epigenetic biomarkers for general processes, such as inflammation or neoplasia, could be of use in the diagnosis and clinical management of IBD — for example distinguishing IBD from other causes of gastrointestinal symptoms, monitoring disease activity and response to treatment, and identifying those at imminent risk of neoplasia.

Epigenetic studies of IBD have examined both blood and intestinal tissue, with findings reported in each. While studying blood adds a further abstraction between the sites of disease and the studied tissue, there are reasons to support studying blood cells in IBD — namely the role of blood derived immune cells in gut immunity, extraintestinal manifestations of disease, and the response to autologous stem cell transplantation. The greatest advantage of studying blood is ease of access, with less invasive methods reducing costs and increasing acceptability to patients and controls. The development of clinically useful biomarkers in blood could offer cheaper and more rapid initial diagnostic tests and allow increased testing frequency for monitoring response to treatment, and earlier

detection of exacerbations or neoplastic transformation. Epigenetic biomarkers have proven feasible in small studies, and it would be reasonable to predict the successful development of biomarkers based on DNA methylation or microRNA profiling to accurately diagnose, distinguish phenotypes, and predict outcomes and response to treatment. However, at this stage very little of the necessary validation work for clinical deployment has been undertaken.

Growing classes of therapies based on modulating all major epigenetic mechanisms are either in use or under active development for a range of conditions. While the lack of specificity demonstrated by many disease-associated epigenetic findings is a frequent criticism, this also means that epigenetic treatments modulating core inflammatory pathways could be of benefit to many diseases.

There is as yet insufficient understanding of the effectiveness and permanence of reversing histone marks and DNA methylation and long-term safety data is currently scarce. The ability to deliver endogenous microRNAs or specifically targeted short non-coding RNAs orally, with reasonable bioavailability and half-life, and minimal immune stimulation has great potential.

Chapter 3

Defining the Methylome of Crohn's Disease in Children

3.1 Introduction

Our group had previously published a paper comparing DNA methylation between 21 adult women with ileal Crohn's disease and 19 healthy adult female controls using the Illumina 27k chip, with additional analysis of 16 female children with Crohn's disease [458]. The analysis identified methylation changes affecting known CD pathogenic pathways and susceptibility genes, and other groups have similarly demonstrated multiple methylation changes in the blood and intestine of IBD patients (see table 2.4 on page 33).

The aim of the present study was to perform a similar analysis in a larger group of patients with the recently released Illumina 450k chip allowing much greater coverage of the epigenome. Some paediatric samples had been included in the Nimmo *et al.* analysis [458], but all other published studies had been in adults. We hypothesized that children would be a superior study population for epigenetic work as the confounding effects of age, comorbidity, polypharmacy and environmental exposures would be lessened. However, we also appreciated that if disease risk in children carried a larger genetic component and epigenetic changes were predominantly environment-related then epigenetic differences between cases and controls could be greatly reduced. Finally, appreciating the differences between adult and paediatric IBD (see section 2.6, page 43), some proportion of any findings may be specific to childhood-onset disease, and any results would need to be replicated in adults before they could be interpreted as important to adult disease.

The central hypothesis was that there would be significant and reproducible methylation differences between children with Crohn's and symptomatic controls. There could be many potential causes for any findings — environmental, intrauterine, genetic, shifts in leucocyte proportions, or differential expression. However, any differences would either be correlated with products and pathways involved in the pathogenesis of Crohn's, or disrupted by the effects of Crohn's. Identifying these differences may identify novel genes of interest.

The previous work by Nimmo *et al.* found that methylation differences were enriched in proximity to the 71 then identified GWAS risk loci. We hypothesised that methylation differences in children would also be enriched in proximity to GWAS risk loci, and that this may provide more granular information about the large regions found to carry a genetic risk by GWAS.

The first cohort, consisting of children who were investigated for suspected IBD and either confirmed to have Crohn's disease or found to be healthy matches the population in whom a diagnostic biomarker would be used. We hypothesized that methylation differences between the two groups would be sufficient to accurately distinguish between children with Crohn's disease and symptomatic but healthy controls, and perhaps allow stratification of Crohn's patients.

Some sections of this chapter have been published in:

Adams AT, *et al.* Two-stage Genome-wide Methylation Profiling in Childhood-onset Crohn's Disease Implicates Epigenetic Alterations at the VMP1/MIR21 and HLA Loci. *Inflammatory Bowel Diseases.* 2014;20(10):1784–93.

3.2 Materials & methods

3.2.1 Patient recruitment and selection

Paediatric patients Samples were collected in centres across Scotland. The Bacteria in IBD in Scottish Children Undergoing Investigation before Treatment (BISCUIT) study provided peripheral blood leukocyte DNA for the discovery cohort from 18 treatment-naive newly diagnosed patients and 18 age and sex-matched non-diseased controls from Aberdeen, Glasgow, and Dundee. Controls were selected from the "normal colon control" subgroup of the BISCUIT study who had microscopically and macroscopically normal colons, as opposed to the eosinophilic control, non-specific inflamed non-IBD, proto-IBD, or disease groups. These children had been rigorously investigated for gastrointestinal symptoms indicative of IBD but did not have or subsequently develop any organic gastrointestinal pathology, including IBD [244, 243]. There was no significant difference between this control group and children with IBD in rates of abdominal pain, tenesmus, blood in stool, constipation, nausea, vomiting, heartburn, weight loss or poor growth; however, there were reduced rates of anorexia (33.3 vs 70.5%, p=0.001), diarrhoea (59.5 vs 84.1%, p=0.016), and blood when wiping (40.5 vs 65.9%, p=0.03).

There was also no significant difference between the children found to have IBD and those normal colon controls in rates of previous surgery (gastrointestinal or otherwise), comorbid atopic conditions, previous use of antibiotics, steroids, or acid suppression, exposure to smoking or pets in the home, vaginal delivery and breastfeeding. The normal colon control children did have a lower rate of concurrent upper GI endoscopy (83.3 vs 100%, p=0.005), and a lower rate of histological gastritis on upper GI endoscopy (34.3 vs 86.4%, p=<0.001), but had a higher rate of *H. pylori* on histology from upper GI endoscopy (11.4 vs 0%, p=0.035) [243].

The replication cohort comprised DNA samples from 18 children with established CD supplied by the Paediatric-onset IBD Cohort and Treatment Study (PICTS), [653] analysed against a second set of 18 normal colon controls from the BISCUIT study.

Within both cohorts, patients and controls were matched for age and gender. The BISCUIT study was approved by the North of Scotland Research Ethics Committee (09/S0802/24) and PICTS by ethics committees at participating centres (Edinburgh, Glasgow, Aberdeen, and Dundee, LREC 2002/6/18). Written

informed consent was obtained from the parents of all participating children and informed assent was also obtained from older children capable of understanding the nature of the study. Summary demographics for both cohorts are in shown in table 3.1.

	Discovery Cohort		Replication Cohort	
	Crohn's	Control	Crohn's	Control
Female (%)	29	29	22	22
Median Age	12.2	12.7	11.7	11.3
Age range	7.6–15.5	8.4–15.3	5.8–15.4	6.3–15.4
Neutrophils $\times 10^9$/L	5.0	2.9	4.7	4.0
Lymphocytes $\times 10^9$/L	1.6	2.2	1.8	1.8

Table 3.1: Age, sex and mean neutrophil and lymphocyte counts for the two paediatric cohorts

Adult pyrosequencing replication cohort Adult replication was carried out in samples from patients recruited from gastroenterology clinics at Edinburgh's Western General Hospital for genetic research (LREC 2000/4/192) or into the IBD-BIOM and IBD Character projects. Healthy control samples were obtained from volunteers. Written informed consent was obtained at sample collection from all patients and controls.

	Crohn's	Control
Male:Female	11:9	10:10
Median Age	34.2	34.2
Age Range	20.8–57.1	20.7–57.0
NOD2 WT (%)	70	75
Non-smoker (%)	100	100
White European (%)	100	95

Table 3.2: Summary demographics of the initial adult replication cohort. *NOD2* WT denotes wild type *NOD2* status, defined as not carrying any of the R702W, G908R, or fs1007insC CD-associated variants.

Reanalysis of biopsy microarray data from Noble *et al.* Data from two previous microarray expression studies undertaken by our group [461, 462] were combined and reanalysed to provide preliminary information regarding expression in the gut in response to IBD and inflammation for a small number of genes which contained highly significant CD-associated methylation changes.

Montreal %	-L4	+L4
L1	15	0
L2	55	0
L3	15	5
L4	5	
Oral	5	

Table 3.3: Montreal classification of disease location in the initial adult replication cohort (see table 2.5, page 45 for details)

Participants were recruited from the gastroenterology clinic in the Western General Hospital (Edinburgh, UK), all sample donors gave written informed consent, and Lothian Local Research Ethics Committee approved the study protocol (REC 04/S1103/22). Non-inflamed biopsies were obtained from normal colorectal cancer screening colonoscopies, and investigations of gastrointestinal symptoms which did not reveal any histological abnormality (presumed IBS). Inflamed control biopsies were obtained from patients with a variety of non-IBD diagnoses including pseudomembranous colitis, diverticulitis, amoebiasis, microscopic colitis, eosinophilic infiltration, and scattered lymphoid aggregates. The IBD patients included new diagnoses, quiescent established disease, and active established disease, and provided combinations of inflamed and non-inflamed biopsies from locations between the terminal ileum and rectum.

Paired biopsies were taken from each location, allowing histological charicterisation of inflammation in one, and extraction of RNA from the other. RNA extraction was performed using 'micro total RNA isolation from animal tissues protocol' (Qiagen) and analysed using Agilent Whole Genome microarrays (Agilent), with normalisation against Stratagene Universal Human Reference. Further details can be obtained from the original publications [461, 462], and the raw datasets have been made available online.

3.2.2 Experimental techniques

Bisulfite conversion was performed using EZ DNA Methylation Kit (Zymo Research Corporation, Irvine, CA, USA. Cat no. D5001, D5002).

Genome-wide methylation was analysed using the Infinium HumanMethylation450 BeadChip Kit (Illumina, San Diego, CA, USA), hereafter Illumina 450k. The samples were prepared and run by the Wellcome Trust Clinical Research Facility (Western General Hospital, Edinburgh, UK).

Pyrosequencing was performed using a PyroMark Q24 System (Qiagen) Primers (Table 3.4) were designed using PyroMark Assay Design Software (version 2.0.1.15, Qiagen) and supplied by Sigma-Aldrich (St. Louis). Data was exported from the manufacturers software and analysed in R, with significance determined by Wilcoxon rank-sum test.

Samples were run in duplicate, results with any quality warnings were discarded and repeated, analysis was performed on the mean of two methylation readings within 2% absolute methylation. Where the variance between repeats was greater than 2% a further set of duplicate reactions was assayed, if these two readings and one of the original readings were all within 2% (probably due to a pipetting error or contamination of one of the original samples) the mean of these three was used, otherwise the sample was excluded.

Further explanation of pyrosequencing and the factors involved in sequence quality assessment can be found in section 3.6.1 (page 170). Details of the cohorts used for pyrosequencing are in section 3.2.1.

Gene		Primer
VMP1	F:	TGGGTTATTGTATTTTGTTTTTAGTGTTG
	R:	[Btn]-ACTAACAAACCAACTTCACTTATTTAC
	S:	GTATTTTGTTTTTAGTGTTGTT
RPS6KA2	F:	GGTGGAGTTTATTGGAAGGTTGTG
	R:	[Btn]-ACAAAATCCCTCTAAATCCAACTATCT
	S:	TGGGTGGTTTATTTAGAAT
ARHGEF3	F:	ATAGTTTTGAAGGAAAGATAGTTATGAAAT
	R:	[Btn]-CCAAATTCTCACTTTCCAAACCT
	S:	GAAGGAAAGATAGTTATGAAATAGT
SBNO2	F:	[Btn]-AGGAAAGAAGTTAGGGTTTGAT
	R:	ACTCAATTACCCTCTCCTTTTTT
	S:	CCTAAAAAACTAAATCACCAT
CFI	F:	TTGAAAGAATTTGGTTGAAATTTAGAGAG
	R:	[Btn]-TTCCTATTAATCACCTCCCTCAACT
	S:	TGGTTGAAATTTAGAGAGAT
CDC42BPB	F:	[Btn]-TGTAAAAATGGGTATAGTTTAGTTATGAT
	R:	ATTCCCCTATAACAAACTTCTCA
	S:	AAACACATACCAAACAAAAAAC
SOCS3	F:	TAGTTGGGTGATTTTTTTATAGGAGTT
	R:	[Btn]-CCCCCAAAAAAACCTATTACATCTACT
	S:	AGAGATGTTGAAGAGTG

Table 3.4: Forward (F), reverse (R), and sequencing (S) primers. [Btn]: 5′ biotinylation, 0.05µmol synthesis scale and HPLC purification

3.2.3 Data analysis

Methylation analysis

No samples met the QC exclusion criteria of 1% of probes having a detection p value >0.05, but one sample was removed due to a mismatch between the recorded sex, and the predicted sex based on median intensities of all sex chromosome probes (illustration of this analysis is shown in figure A.1, page 256). Probes were removed from further analysis if they had a beadcount <3 in 5% of samples, if >1% of samples had a detection p value >0.05 for that probe, or if there was a SNP with MAF $\geq$ 0.01 in European samples from the Thousand Genomes Project within 2bp of the probe target, or if they targeted the sex chromosomes, resulting in a final probe count of 449211. Data were corrected using background removal and quantile normalization in the 'lumi' R package [156, 406] followed by beta mixture quantile dilation [612]. Batch effects from clinical centre, Illumina 450k slide and position, and for the combined analysis the cohort, were controlled for using ComBAT [365].

Differential leukocyte counts from the same day that DNA samples were taken were available for 24 patients and 19 controls. Linear models were created for all Illumina 450k probes and disease status in these samples. Probes were selected, which had F test with P values $<10^{-8}$ but a P value for disease association of >0.05. Combinations of 100 random probes meeting these criteria were tested, and the best performing probe set was used to predict the differential cell counts for samples without measured differential leukocyte counts. This model is similar to that described by Houseman *et al.* [270].

Analysis of the methylation status of cases versus controls was performed using limma [582] in R using linear modelling of beta values with diagnosis, age, sex, and neutrophil, other granulocyte, lymphocyte, and monocyte proportions as covariates. The Benjamini-Hochberg false discovery rate (FDR) [52] was calculated for each probe, with a FDR corrected P <0.05 used to define significance in analysis of broader methylation patterns, such as identifying differentially methylated regions. For significance of individual probes, the more conservative Bonferroni correction was used.

Identification of differentially methylated regions

Analysis was performed in R using an algorithm based on the probe lasso algorithm in the ChAMP pipeline [443]. DMRs were defined as 3 consecutive probes

with FDR adjusted p values <0.05 with a shared direction of change in methylation within a distance threshold based upon average probe density in similar genomic regions (Promoter region, 5′ UTR, gene body, *etc*). Further details of the probe lasso algorithm are shown in section A.3.5 (page 262)

Analysis of GWAS correlation

Data analysis was performed in R. Statistical significance of correlations was calculated from the Spearman's rank correlation coefficient. SNPs included in the initial comparison were those from Jostins *et al.* [297] and Liu *et al.* [384] associated with CD or IBD, but excluding those only associated with UC.

For the comparison with GWAS SNPs for non-IBD conditions the NHGRI-EBI Catalog of published genome-wide association studies [264] was obtained using the R package 'gwascat' [97].

Comparisons were made based upon the smallest methylation p value within bins of a defined size, with the analysis repeated for bin sizes of 25kb, 50kb, 100kb, 250kb, 500kb, 1Mb, 2Mb, and 4Mb. Bins were created centred on each GWAS SNP, and 1000 randomly selected control bins of each size were chosen to match the probe density found in GWAS SNP-containing bins.

GO term analysis

The R package 'Goseq' [723] was used as described by Geeleher *et al.* [207] to create a probability weighting function to correct for bias introduced by variation in the number of Illumina 450k probes per gene. Differential methylation for a gene was defined as the presence of a FDR significant probe from the combined paediatric analysis annotated to that gene. P values for GO term enrichment were corrected for multiple testing with Benjamini-Hochberg false discovery rate [52].

Transcription factor consensus sites

2kb genomic regions were defined centred on each Illumina 450k probe with a Bonferroni corrected p<0.05 in the combined analysis. Overlapping regions were combined into the one containing the most significant methylation result. Sequences were obtained for genomic coordinates using the R packages 'BSgenome' and 'BSgenome.Hsapiens.UCSC.hg19' [477, 616].

The R package 'BCRANK' [11] was used to search for DNA motifs in the defined regions by starting with a random seed and iterating minor alterations

to find a local optimum for frequency. This was repeated with 1000 random starting seeds, with each local optimum ranked by its number of occurrences with a weighting toward occurrences in sequences with more significant methylation changes and penalties for redundant bases (IUPAC codes other than A,C,G, or T) and being found in repetitive elements.

A normalised position weight matrix (PWM) was produced for each motif and these were compared to experimentally derived PWMs for transcription factors in the JASPAR database (2016 version) [415]. Non-human results were ignored and where there was overlap between results, the strongest is presented.

The analysis was repeated twice using Illumina 450k probes with FDR adjusted $p<0.05$, and the most significant probe within each DMR as the centre of each region. The strongest results from each set of sequences were used as starting seeds for a further analysis in the other groups of sequences to ensure that they had not been missed by the random seeds.

Power calculation

Power calculations were performed for two-tailed T tests using the difference in means relative to the population standard deviation for each for the Bonferroni significant probes for a power of 80% and a significance level of 1.1×10^{-7}, equivalent to a Bonferroni corrected $p<0.05$, for 449211 probes.

$$\delta = \frac{\bar{\beta}_{CD} - \bar{\beta}_{HC}}{\sqrt{(\sigma^2_{CD} + \sigma^2_{HC})/2}}$$

CpG Island detection at significant probes

Sequences were extracted and combined as described in section 3.2.3. CpG islands were defined as 200bp spans containing $>0.6\times$ the expected number of CpGs in a random sequence (7.5 per 200bp) with a GC content >50%. 20 sets of 1000 non-significant probes were randomly selected for controls.

3.3 Results

3.3.1 Combining the two paediatric cohorts

Analysis of the discovery cohort found 9 probes with FDR adjusted p <0.05, of which only the least significant result failed to be replicated in the second cohort (see table 3.5).

			P Value		
Gene	Chr	Coord	Discovery	Replication	Combined
VMP1	chr17	57915717	2.19×10^{-10}	1.55×10^{-5}	1.97×10^{-15}
VMP1	chr17	57915665	9.11×10^{-10}	9.25×10^{-5}	2.65×10^{-13}
	chr3	101901234	3.82×10^{-9}	5.14×10^{-5}	1.59×10^{-13}
RPS6KA2	chr6	166970252	1.17×10^{-8}	3.73×10^{-6}	4.47×10^{-15}
ARHGEF3	chr3	57041402	3.32×10^{-8}	4.69×10^{-5}	1.44×10^{-12}
	chr22	50327986	1.71×10^{-7}	3.46×10^{-3}	3.00×10^{-9}
	chr6	35696870	4.72×10^{-7}	2.08×10^{-2}	3.05×10^{-7}
MYO1E	chr15	59588622	4.86×10^{-7}	1.71×10^{-2}	4.07×10^{-9}
GGT6	chr17	4464400	7.55×10^{-7}	1.02×10^{-1}	6.9×10^{-7}

Table 3.5: Individually significant CpGs from the discovery cohort (Benjamini-Hochberg FDR<0.05) which replicated in the second cohort (P<0.05) using recalculated P values including predicted cell counts. All P values shown uncorrected, ranked in order of significance in the discovery cohort

Overall there was a high degree of correlation in the results between the two cohorts, with 97.3% of probes with nominally significant disease-associated methylation changes in each cohort show the same direction of change (see figure 3.1).

The two cohorts were therefore analysed as a combined cohort, a summary of the number of probes with disease associated methylation changes is shown in table 3.6, a Manhattan plot of the combined paediatric Illumina 450k data is shown in figure 3.2 (page 61), and a QQ plot in figure 3.3 (page 62).

Significance threshold	Number of probes	Percent of total
p<0.05	44062	9.8%
FDR<0.05	1319	0.29%
Bonferroni<0.05	65	0.01%

Table 3.6: The number of probes with methylation differences between Crohn's disease and symptomatic controls in the combined analysis by significance threshold.

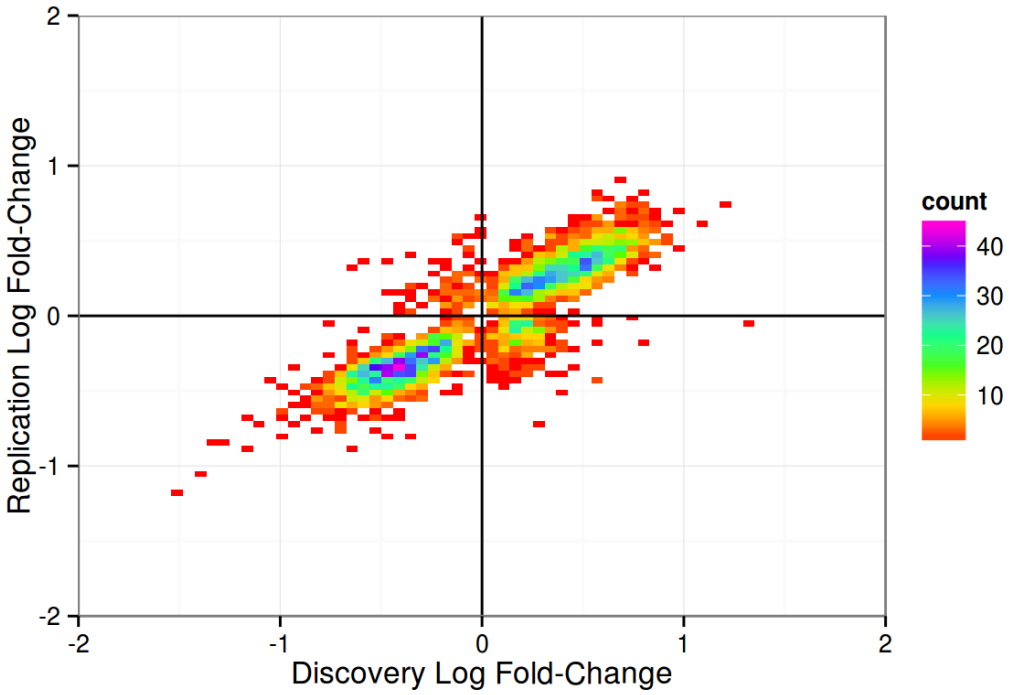


Figure 3.1: Correlation in direction of change between paediatric cohorts. All probes with nominal significance in both cohorts included (n=3620), correlation between log fold-change in the two cohorts for these probes: ρ=0.87, p<2.2×10^{-308}.

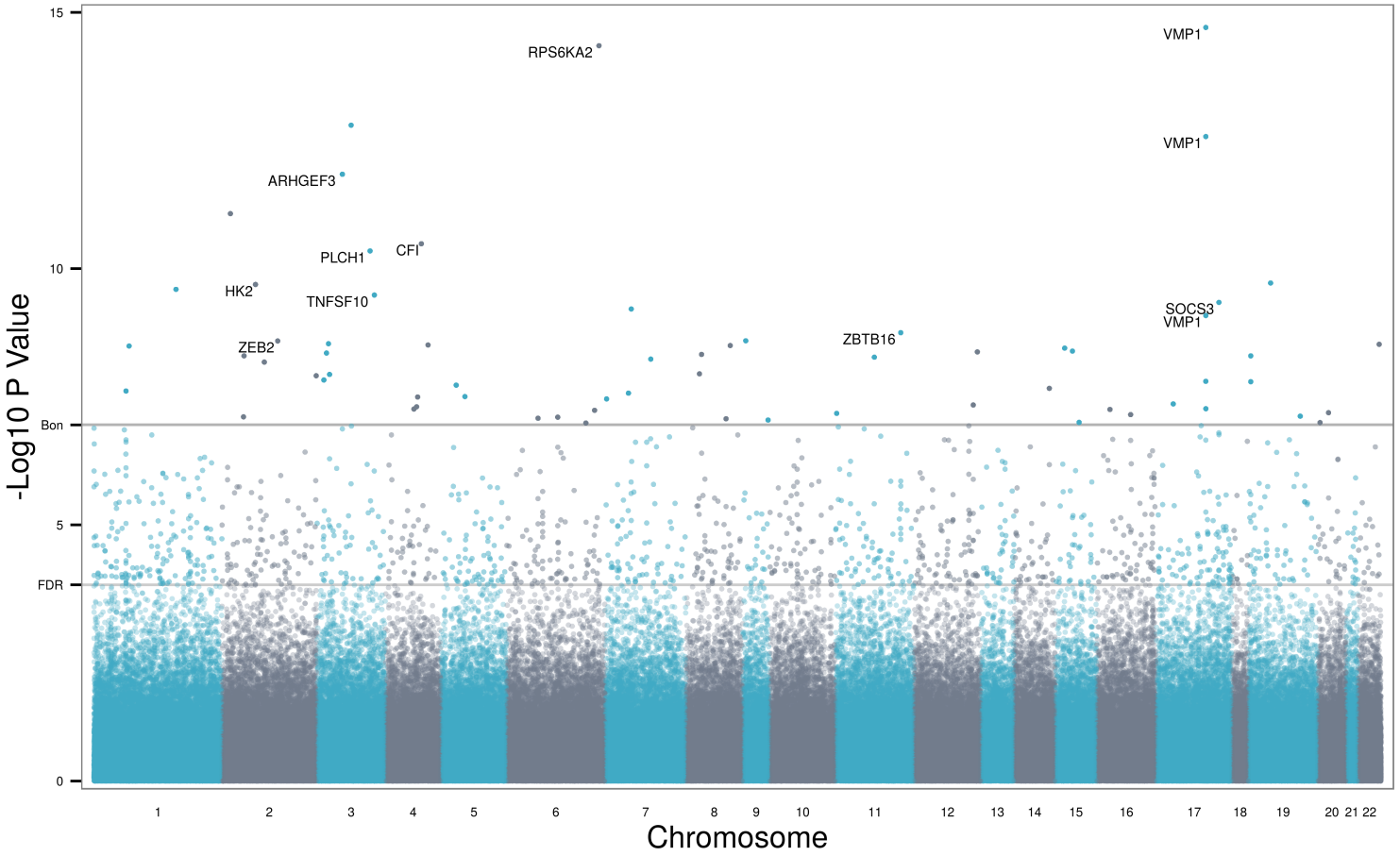


Figure 3.2: A Manhattan plot of the combined paediatric Illumina 450k data. Horizontal lines indicate Bonferroni and FDR significance thresholds. Probes within the top 20 results which are annotated to a gene are labelled.

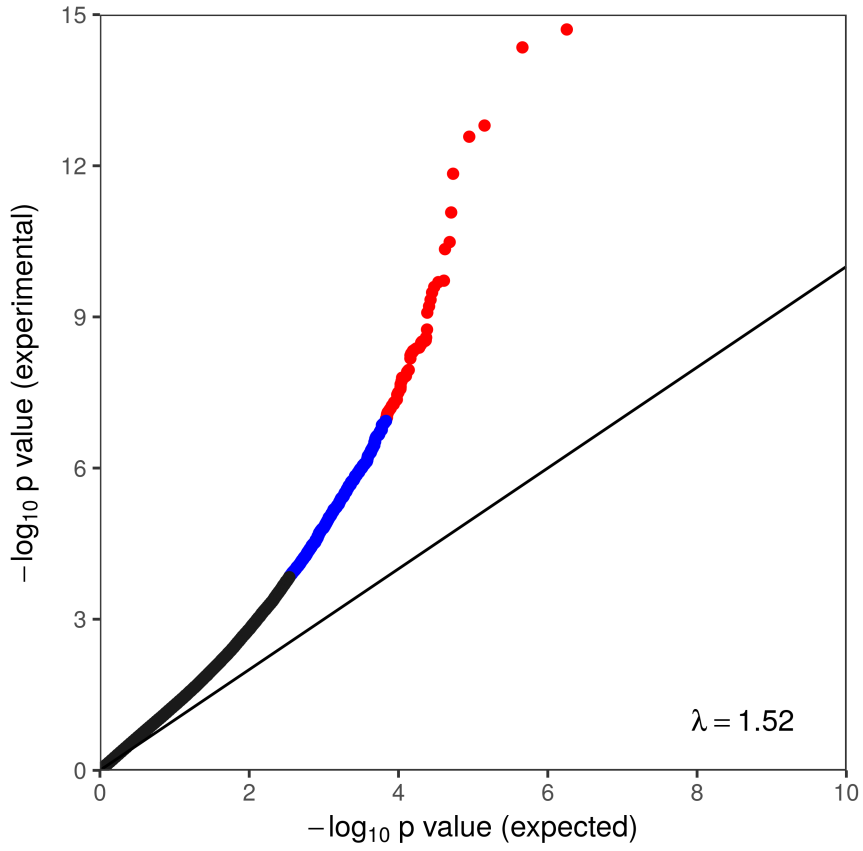


Figure 3.3: QQ plot of the combined paediatric Illumina 450k data. Bonferroni significant probes are shown in red, FDR significant probes are shown in blue. The inflation factor (λ) is shown, and the solid line shows the position of λ=1.

3.3.2 Methylation profile

There was no significant difference in overall mean methylation between the CD and control samples (p=0.24), however both FDR (n=1319, β_{CD}-β_{HC}=-0.006, T test p=0.0005) and Bonferroni (n=65, β_{CD}-β_{HC}=-0.043, p=1.14×10^{-14}) significant probes showed overall hypomethylation compared to controls.

The methylation probes showed a bimodal distribution with peaks at 12.6 and 83.0% methylation. The peaks for CD were shifted further from the centre by 0.23% and 0.7% (T test p=1.05×10^{-10} and 1.80×10^{-147}) for the low and high peaks respectively. The mean methylation at significant probes is markedly different to non-significant probes, with a large enrichment of probes with intermediate methylation (see figure 3.4).

Overall, more probes were hypermethylated in CD (56.7% hypermethylated, p<2.22×10^{-308}), however FDR and Bonferroni significant probes were more likely to be hypomethylated (54.5%, p=0.001 & 76.9%, p=1.42×10^{-5}). Figure 3.5 (page 65) shows a volcano plot of the paediatric Illumina 450k data, demonstrating the increased prevalence of hypomethylation among the most significant probes.

Overall, hypermethylated probes had a slightly larger difference in means between CD and controls (β_{CD}-β_{HC}=0.0002, χ^2 Test p=1.86×10^{-14}), but in FDR and Bonferroni significant probes the hypomethylated probes showed a greater difference between CD and controls (β_{CD}-β_{HC}=-0.007, p=2.77×10^{-16} and β_{CD}-β_{HC}=-0.022, p=1.0×10^{-4}).

In 63% of Bonferroni significant probes, the range of methylation values was larger for CD patients than controls, compared to 41% in FDR significant probes and 44% in all probes. In the Bonferroni significant probes where the range was larger in CD the difference was up to 0.29 (mean 0.076), in the probes where the range was smaller in CD the difference was up to 0.079 (mean 0.027). In all probes there was a small (Pearson's product moment correlation coefficient r=-0.06) but significant (p<2.22×10^{-308}) correlation between a negative β difference (*i.e.* hypomethylation in CD) and increased range of methylation results in CD compared to controls. This correlation was stronger in FDR significant probes (r=-0.41, p=4.02×10^{-57}) and strongest in Bonferroni significant probes (r=-0.68, p=6.84×10^{-10}).

Figure 3.6 (page 66) shows probability densities for methylation at Bonferroni significant probes and figure 3.7 (page 67) shows the β difference and β range difference in all probes. The general pattern described above of hypomethylation and increased variance in significant probes can be seen clearly.

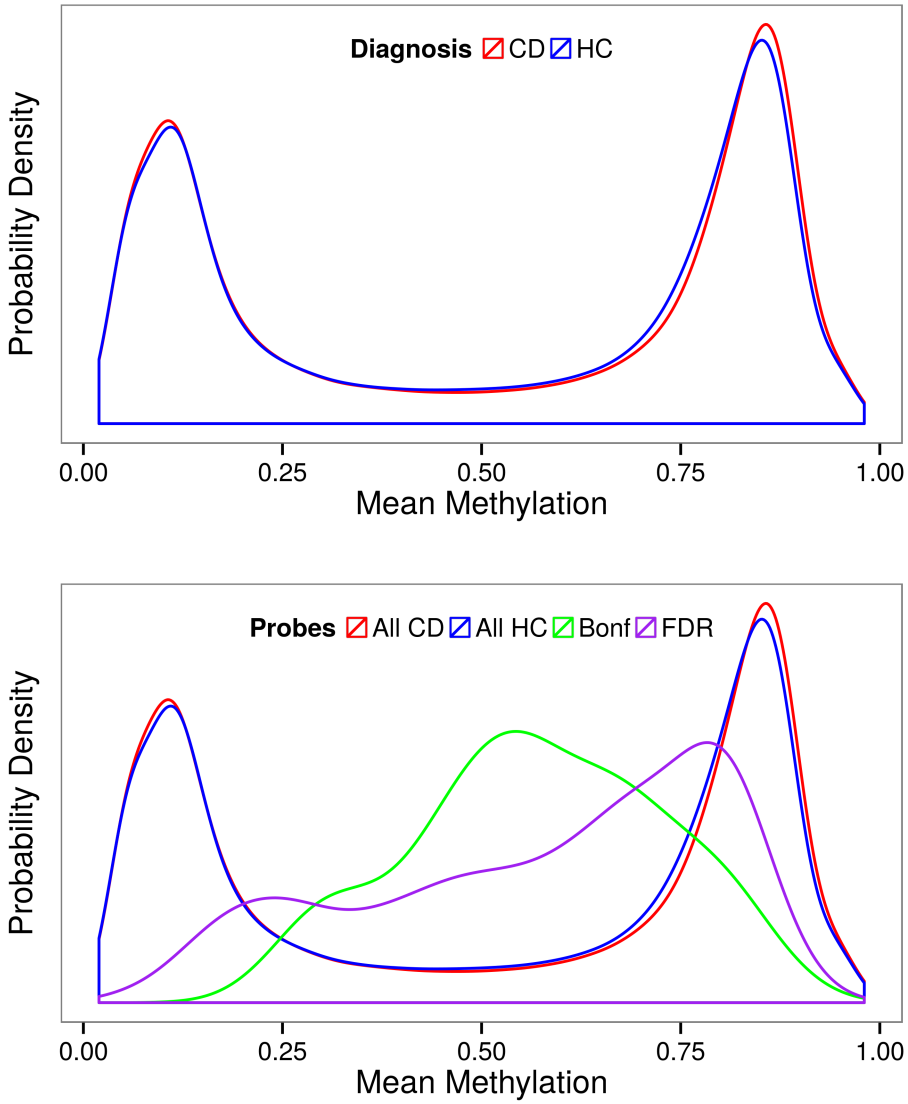


Figure 3.4: Mean methylation results for Crohn's and control, with Bonferroni and FDR significant probes superimposed below

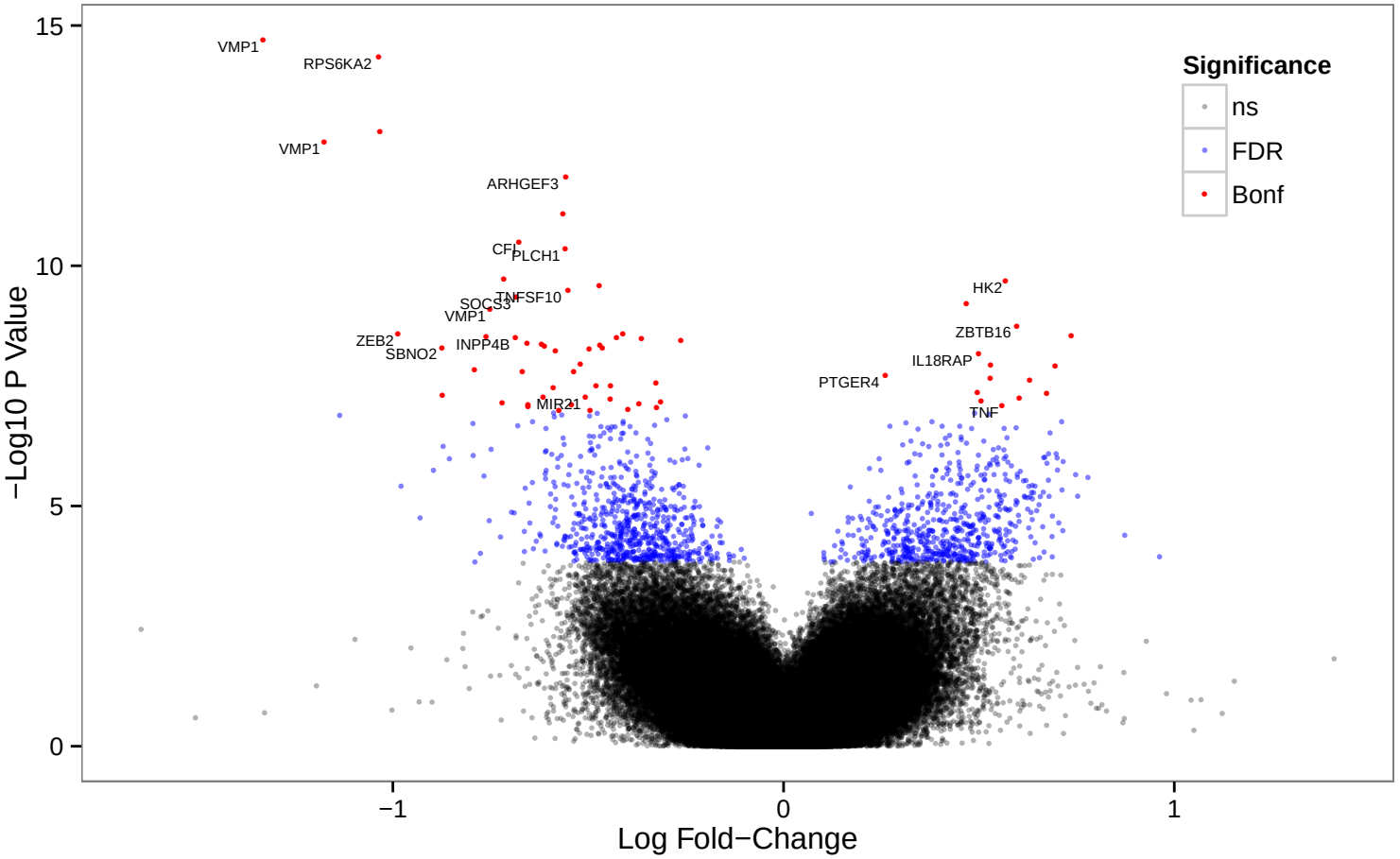


Figure 3.5: A volcano plot of the combined paediatric Illumina 450k data

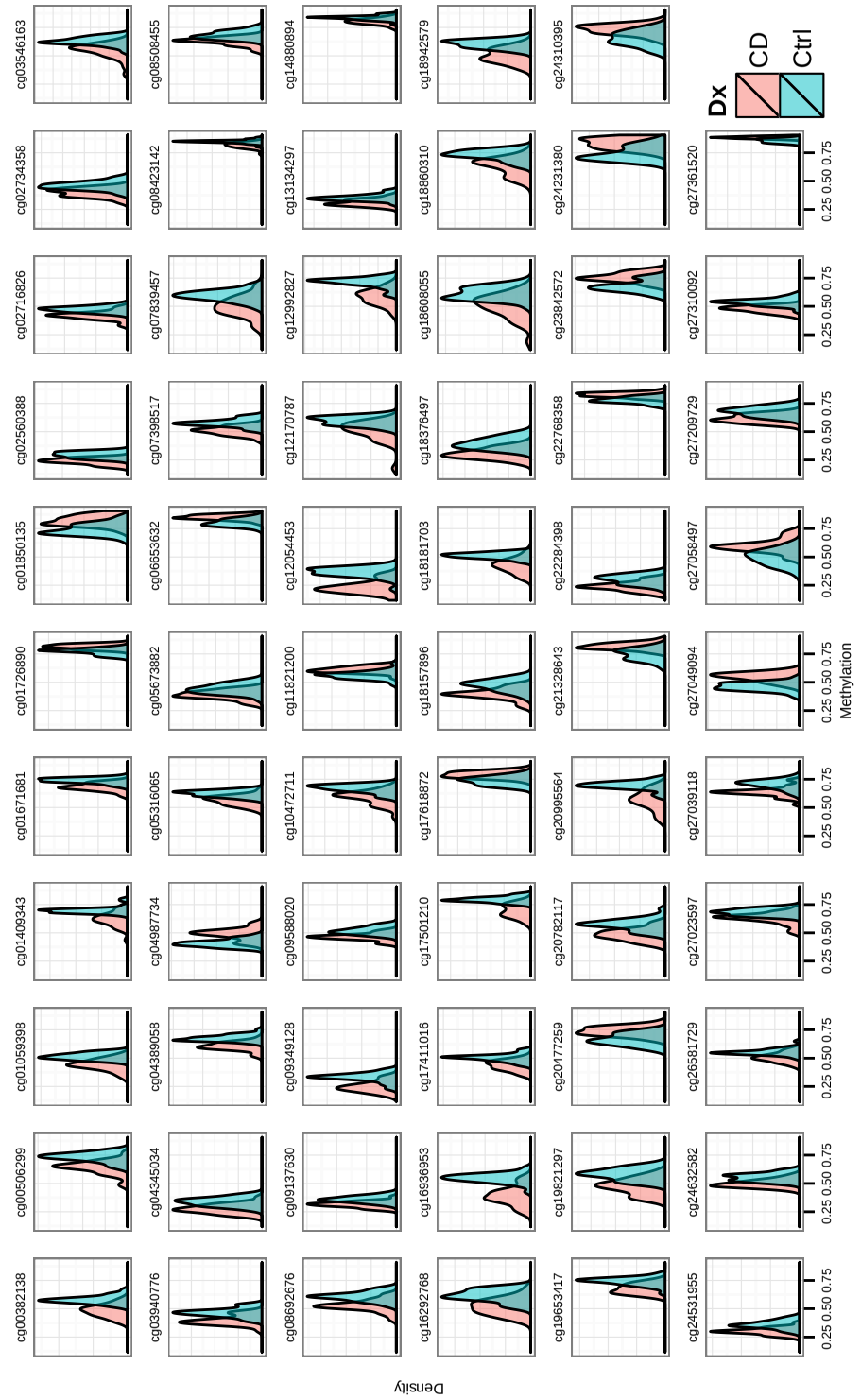


Figure 3.6: Probability densities for methylation values at each of the 65 Illumina 450k probes with Bonferroni significant methylation changes

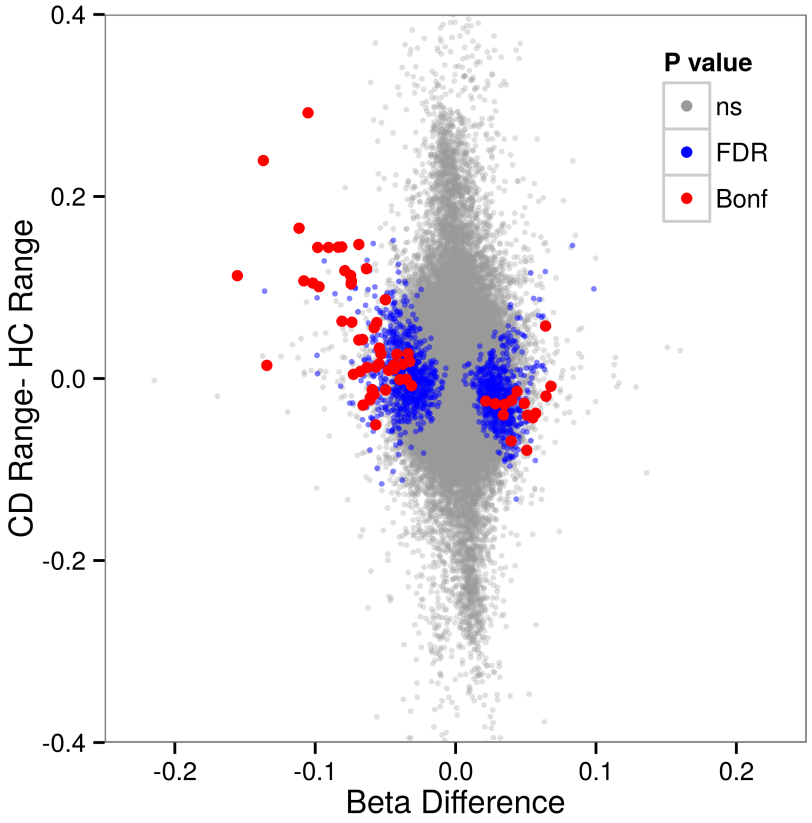


Figure 3.7: The most significant probes are predominantly hypomethylated in CD and show increased variability in CD

			Group Size		
Gene	β_{CD}-β_{HC}	σ_{pop}	$p<0.05$	$p<1.5\times10^{-4}$	$p<1.1\times10^{-7}$
VMP1†	-0.134	0.050	3	10	17
RPS6KA2	-0.099	0.045	4	12	22
SOCS3	-0.089	0.048	6	16	28
SBNO2	-0.114	0.083	9	26	47
RUNX3	-0.053	0.064	7	21	38
POLK‡	-0.103	0.049	10	29	52

Table 3.7: The group sizes needed for 80% power to detect methylation changes equivalent to representative significant findings from this study including the smallest (†) and largest (‡) results, together with the β difference and population standard deviation. Significance thresholds shown are for nominally significant, an estimation for FDR significance, and Bonferroni significance based on the number of probes in the Illumina 450k platform.

Power calculations were performed using the delta value for each of the Bonferroni significant probes, and the group sizes required to detect differences matching a selection of these probes at 80% power are shown in table 3.7. Figure 3.8 shows the group sizes necessary for every probe with power between 50–95% at a significance level of 1.1×10^{-7}.

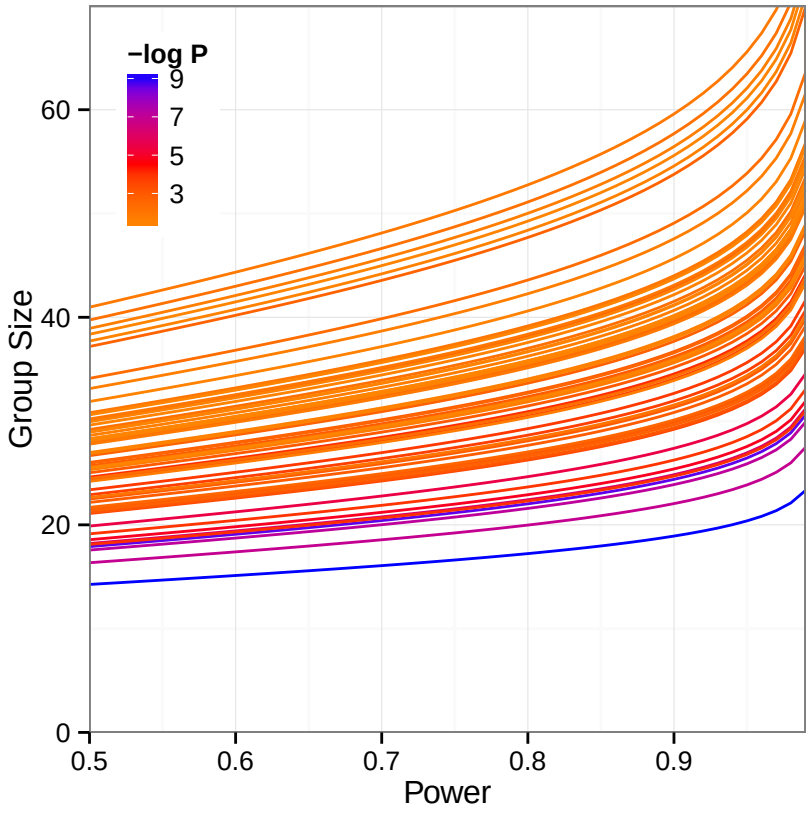


Figure 3.8: The group sizes required to detect methylation differences matching the 65 Bonferroni significant results at significance levels equivalent to Bonferroni corrected significance in the Illumina 450k platform. Coloured by Bonferroni corrected p value.

3.3.3 Individual CpGs

An initial analysis of the combined paediatric data used neutrophil:lymphocyte ratios as a covariate rather than the procedure described in section 3.2.3. These data were presented in abstracts to the British Society of Gastroenterology [2] and the European Crohn's and Colitis Organisation [313] (see B.1 and B.2, page 289). In this analysis there were 165 significant probes after Bonferroni correction for multiple testing versus the 65 probes (table 3.9, page 73) found with the reimplemented Houseman method in the final analysis. All significant probes in the final analysis were also significant in the preliminary analysis.

A permutation test (n=1000) was performed by randomly permuting the diagnoses and beta values of each sample and repeating the analysis. A p value cutoff of $p<2.0\times10^{-7}$ would exclude the lowest p value from 95% the random permutations. The uncorrected p value equivalent to a Bonferroni corrected $p<0.05$ is 1.1×10^{-7}, and the p value cutoff of 2.0×10^{-7} would class 82 CpGs as significant — a 26% increase.

Bonferroni significant probes were examined to see if they were in CpG islands. After combining neighbouring significant probes, 19/61 (31.1%) were in CpG islands. This is significantly lower (χ^2 $p=4.58\times10^{-6}$) than the rate observed in 20 random selections of 1000 non-significant probes (60.7%, 95% CI 60.0–61.4%). FDR significant probes were also significantly (19.4%, $p=1.06\times10^{-166}$) less likely to be in CpG islands than the randomly selected non-significant probes. Figure 3.9 shows CpG prevalence, GC content and CGIs in proximity to Bonferroni significant probes.

Both FDR and Bonferroni significant probes showed a positive correlation between the β difference in Crohn's disease and the probability of a probe being in a CGI (FDR: r=0.20, $p=1.47\times10^{-11}$, Bonferroni: r=0.38, p=0.003, Pearson's product-moment correlation). Probes which were not in CGIs were more likely to be hypomethylated (see table 3.8). There was no correlation between the p value for the methylation change and the chance of a probe being in a CGI.

A

		Meth ↓	Meth ↑
CGI	N	36	6
	Y	10	9

B

	Meth ↓	Meth ↑
N	574	348
Y	87	135

Table 3.8: Hypomethylation correlates with not being in a CGI for (A) Bonferroni significant regions and (B) FDR significant regions χ^2 p=0.009 and 1.00×10^{-5}

				Uncorrected P Value			Bonferroni
Gene	Chr	Coord	LogFC	Discovery	Replication	Combined	Combined
VMP1	chr17	57915717	-1.33	2.19×10^{-10}	1.55×10^{-5}	1.97×10^{-15}	8.86×10^{-10}
RPS6KA2	chr6	166970252	-1.04	1.17×10^{-8}	3.73×10^{-6}	4.47×10^{-15}	2.01×10^{-9}
	chr3	101901234	-1.03	3.82×10^{-9}	5.14×10^{-5}	1.59×10^{-13}	7.13×10^{-8}
VMP1	chr17	57915665	-1.18	9.11×10^{-10}	9.25×10^{-5}	2.65×10^{-13}	1.19×10^{-7}
ARHGEF3	chr3	57041402	-0.56	3.32×10^{-8}	4.69×10^{-5}	1.44×10^{-12}	6.48×10^{-7}
	chr2	11969958	-0.56	1.93×10^{-6}	1.60×10^{-5}	8.45×10^{-12}	3.80×10^{-6}
CFI	chr4	110723299	-0.68	1.88×10^{-6}	2.60×10^{-5}	3.27×10^{-11}	1.47×10^{-5}
PLCH1	chr3	155421735	-0.56	4.43×10^{-6}	5.23×10^{-5}	4.52×10^{-11}	2.03×10^{-5}
	chr19	12890029	-0.72	3.35×10^{-6}	7.09×10^{-5}	1.91×10^{-10}	8.60×10^{-5}
HK2	chr2	75067716	0.57	9.03×10^{-6}	8.62×10^{-6}	2.04×10^{-10}	9.18×10^{-5}
	chr1	151945663	-0.47	2.67×10^{-6}	2.56×10^{-5}	2.54×10^{-10}	1.14×10^{-4}
TNFSF10	chr3	172235808	-0.55	9.77×10^{-6}	1.63×10^{-4}	3.28×10^{-10}	1.47×10^{-4}
SOCS3	chr17	76354621	-0.68	3.05×10^{-5}	2.43×10^{-5}	4.56×10^{-10}	2.05×10^{-4}
	chr7	38370874	0.47	2.22×10^{-5}	1.50×10^{-4}	6.14×10^{-10}	2.76×10^{-4}
VMP1	chr17	57915773	-0.75	1.69×10^{-5}	2.99×10^{-4}	8.23×10^{-10}	3.70×10^{-4}
ZBTB16	chr11	113947148	0.60	5.09×10^{-5}	8.38×10^{-5}	1.78×10^{-9}	8.01×10^{-4}
	chr9	33447032	-0.41	3.33×10^{-6}	2.56×10^{-3}	2.56×10^{-9}	1.15×10^{-3}
ZEB2	chr2	145172035	-0.99	2.88×10^{-6}	5.73×10^{-4}	2.58×10^{-9}	1.16×10^{-3}
	chr3	37258149	0.74	7.15×10^{-6}	2.09×10^{-4}	2.92×10^{-9}	1.31×10^{-3}
	chr22	50327986	-0.76	1.71×10^{-7}	3.46×10^{-3}	3.00×10^{-9}	1.35×10^{-3}
INPP4B	chr4	143488622	-0.69	7.92×10^{-5}	9.35×10^{-5}	3.09×10^{-9}	1.39×10^{-3}
GSDMC	chr8	130799007	-0.43	4.65×10^{-5}	9.23×10^{-5}	3.17×10^{-9}	1.43×10^{-3}
WASF2	chr1	27755803	-0.36	3.64×10^{-4}	6.80×10^{-6}	3.24×10^{-9}	1.46×10^{-3}
	chr15	41233701	-0.26	2.60×10^{-5}	1.78×10^{-3}	3.56×10^{-9}	1.60×10^{-3}
MYO1E	chr15	59588622	-0.66	4.86×10^{-7}	1.71×10^{-2}	4.07×10^{-9}	1.83×10^{-3}

				Uncorrected P Value			Bonferroni
Gene	Chr	Coord	LogFC	Discovery	Replication	Combined	Combined
	chr12	132654924	-0.62	1.16×10^{-6}	5.05×10^{-3}	4.22×10^{-9}	1.90×10^{-3}
	chr3	30327579	-0.47	3.99×10^{-6}	2.74×10^{-3}	4.44×10^{-9}	2.00×10^{-3}
CLU	chr8	27467783	-0.61	2.45×10^{-5}	3.17×10^{-4}	4.72×10^{-9}	2.12×10^{-3}
SBNO2	chr19	1130866	-0.87	3.63×10^{-5}	4.08×10^{-4}	5.04×10^{-9}	2.26×10^{-3}
	chr2	47100912	-0.46	4.53×10^{-4}	9.78×10^{-6}	5.04×10^{-9}	2.26×10^{-3}
NRXN2	chr11	64428925	-0.50	2.89×10^{-5}	2.17×10^{-4}	5.35×10^{-9}	2.40×10^{-3}
SLC25A13	chr7	95865631	-0.58	1.46×10^{-4}	2.14×10^{-5}	5.83×10^{-9}	2.62×10^{-3}
IL18RAP	chr2	103038171	0.50	3.35×10^{-5}	3.04×10^{-3}	6.67×10^{-9}	3.00×10^{-3}
LOXL2	chr8	23154691	-0.52	7.67×10^{-4}	1.92×10^{-5}	1.13×10^{-8}	5.09×10^{-3}
CX3CR1	chr3	39309435	0.53	1.87×10^{-4}	1.25×10^{-5}	1.16×10^{-8}	5.21×10^{-3}
RTP5	chr2	242813914	0.70	6.21×10^{-4}	3.57×10^{-5}	1.23×10^{-8}	5.51×10^{-3}
RFTN1	chr3	16469127	-0.79	6.15×10^{-5}	2.84×10^{-4}	1.49×10^{-8}	6.69×10^{-3}
VMP1	chr17	57915740	-0.67	6.28×10^{-6}	2.17×10^{-3}	1.58×10^{-8}	7.08×10^{-3}
SBNO2	chr19	1130965	-0.54	5.85×10^{-5}	6.85×10^{-4}	1.61×10^{-8}	7.24×10^{-3}
PTGER4	chr5	40685989	0.26	4.52×10^{-5}	4.54×10^{-4}	1.88×10^{-8}	8.45×10^{-3}
CDC42BPB	chr14	103415873	0.53	1.20×10^{-4}	1.15×10^{-4}	2.17×10^{-8}	9.73×10^{-3}
RUNX3	chr1	25291546	0.63	1.21×10^{-3}	1.72×10^{-6}	2.45×10^{-8}	1.10×10^{-2}
	chr7	30737556	-0.33	3.11×10^{-5}	6.18×10^{-3}	2.69×10^{-8}	1.21×10^{-2}
POLK	chr5	74862702	-0.44	7.85×10^{-5}	1.36×10^{-3}	3.14×10^{-8}	1.41×10^{-2}
GPRIN3	chr4	90227074	-0.48	2.30×10^{-4}	1.96×10^{-5}	3.20×10^{-8}	1.44×10^{-2}
HEATR2	chr7	797592	-0.59	4.57×10^{-6}	1.16×10^{-3}	3.49×10^{-8}	1.57×10^{-2}
MPRIP	chr17	17030253	0.50	3.39×10^{-5}	6.88×10^{-4}	4.37×10^{-8}	1.96×10^{-2}
SLC15A4	chr12	129281444	0.67	1.94×10^{-3}	6.74×10^{-5}	4.58×10^{-8}	2.06×10^{-2}
SLC10A6	chr4	87752504	-0.87	4.10×10^{-6}	1.59×10^{-3}	4.97×10^{-8}	2.23×10^{-2}
MIR21	chr17	57918262	-0.51	4.48×10^{-5}	1.20×10^{-3}	5.42×10^{-8}	2.44×10^{-2}

				Uncorrected P Value			Bonferroni
Gene	Chr	Coord	LogFC	Discovery	Replication	Combined	Combined
CNOT6L	chr4	78697859	-0.62	3.27×10^{-6}	6.39×10^{-4}	5.49×10^{-8}	2.46×10^{-2}
NMRAL1	chr16	4516078	0.60	1.99×10^{-5}	2.05×10^{-3}	5.60×10^{-8}	2.52×10^{-2}
SYNJ2	chr6	158490013	-0.44	2.65×10^{-5}	1.39×10^{-3}	5.82×10^{-8}	2.61×10^{-2}
CBFA2T2	chr20	32235621	0.51	8.80×10^{-5}	2.58×10^{-3}	6.47×10^{-8}	2.91×10^{-2}
PTDSS2	chr11	470429	-0.31	5.16×10^{-4}	5.31×10^{-5}	6.68×10^{-8}	3.00×10^{-2}
NLRC5	chr16	57023022	-0.72	4.83×10^{-5}	2.96×10^{-3}	7.06×10^{-8}	3.17×10^{-2}
ZC3H4	chr19	47618017	-0.37	2.02×10^{-4}	1.08×10^{-4}	7.57×10^{-8}	3.40×10^{-2}
PRKCE	chr2	46119607	-0.54	4.40×10^{-4}	7.47×10^{-4}	7.81×10^{-8}	3.51×10^{-2}
FKBP5	chr6	35654363	-0.65	1.76×10^{-4}	6.73×10^{-4}	7.93×10^{-8}	3.56×10^{-2}
TNF	chr6	31544960	0.56	3.44×10^{-5}	4.47×10^{-3}	8.31×10^{-8}	3.73×10^{-2}
TRPS1	chr8	116575902	-0.65	1.02×10^{-4}	1.19×10^{-3}	8.51×10^{-8}	3.82×10^{-2}
NPDC1	chr9	139939792	-0.32	5.34×10^{-4}	8.68×10^{-4}	9.01×10^{-8}	4.05×10^{-2}
	chr15	70797389	-0.40	3.67×10^{-4}	2.76×10^{-4}	1.00×10^{-7}	4.49×10^{-2}
	chr20	2795593	-0.50	2.95×10^{-4}	6.11×10^{-4}	1.01×10^{-7}	4.54×10^{-2}
TBPL1	chr6	134299594	-0.58	1.58×10^{-4}	1.83×10^{-4}	1.03×10^{-7}	4.64×10^{-2}

Table 3.9: Details of all Illumina 450k probes attaining Bonferroni corrected p values <0.05 in the combined analysis. Uncorrected p values shown for the individual cohorts, and combined, as well as Bonferroni corrected p values for the combined analysis. LogFC shows Log_2 fold-change in the combined analysis.

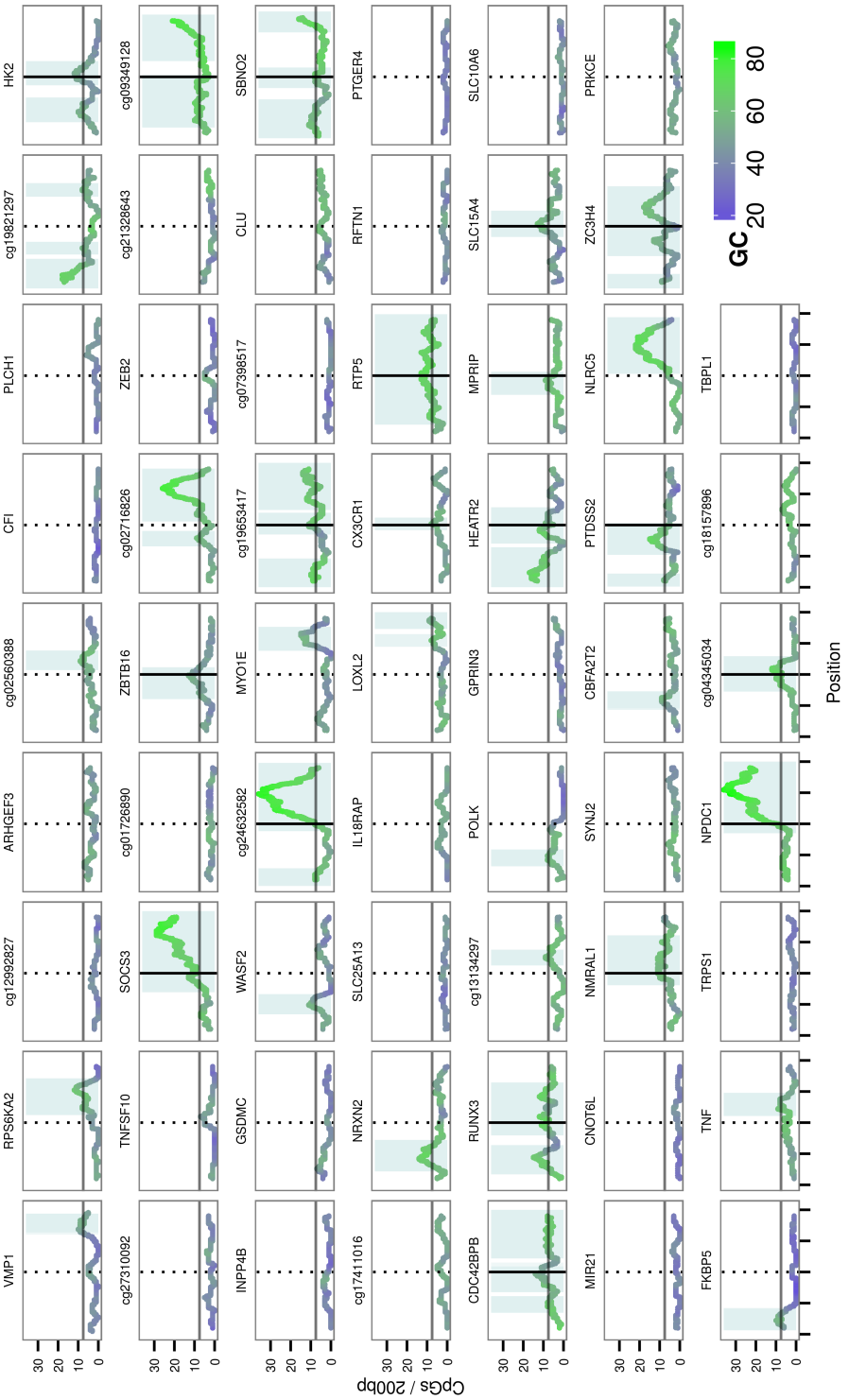


Figure 3.9: Number of CpGs within 200bp and GC content for 1kb regions either side of Bonferroni significant probes. Shaded areas represent CGIs, the horizontal line marks 60% of expected CpGs, the vertical line marks the probe position and is solid if the probe is within a CGI.

3.3.4 Differentially methylated regions

There were 77 probes with FDR adjusted p<0.05 in 19 DMRs of consecutive significant unidirectional methylation change within a distance threshold based on local probe density (table 3.10).

Direction	Gene	P (min)	Chr	Coordinate	Length	Probes
↓	VMP1	1.97×10^{-15}	17	57915150	1138	4
↑	ZBTB16	1.78×10^{-9}	11	113946633	1241	3
↑	SLC15A4	4.58×10^{-8}	12	129281291	408	3
↑	TNF	8.31×10^{-7}	6	31543764	2126	4
↑	DIABLO	1.17×10^{-7}	12	122711835	306	6
↑	ZFYVE28	1.76×10^{-7}	4	2321442	1151	3
↑	RUNX3	2.19×10^{-7}	1	14181794	351	5
↓	HGF	2.44×10^{-7}	7	81399131	536	3
↓	GPR56	6.43×10^{-7}	16	57662273	536	3
↑	TOLLIP	8.18×10^{-7}	11	1296881	554	3
↑	ITGB2	1.21×10^{-6}	21	46341036	536	3
↑	PIEZO1	1.19×10^{-6}	16	88831961	1086	3
↑	PRF1	1.47×10^{-6}	10	72362714	190	6
↑	ZBTB12	2.23×10^{-6}	6	31867623	330	9
↑	PRF1	6.32×10^{-6}	10	72360292	160	6
↑	CD247	7.59×10^{-6}	1	167486463	1685	4
↓	CETP	1.39×10^{-5}	16	56995572	536	3
↑	UBASH3A	1.45×10^{-5}	21	43823747	1030	3
↑	WRAP73	3.21×10^{-5}	1	3563024	1937	3

Table 3.10: Differentially methylated regions. ↑: Hypermethylation, ↓: Hypomethylation, P (min): smallest p value within the DMR.

In contrast to significant probes which were not in differentially methylated regions (see section 3.3.2), DMR probes were significantly more likely to be hypermethylated. Crohn's-associated hypermethylation was seen in 83% compared to 56.7% in all probes, 45.5% in other FDR significant probes (χ^2 Test $p=3.34\times10^{-11}$) and 23.1% in Bonferroni significant probes. Including only one probe per DMR does not change this result (e.g. DMR hypermethylation 79% vs. FDR hypermethylation 45.5%, p=0.001), and there was no significant difference in the number of probes in hypomethylated versus hypermethylated DMRs (3.25 vs 4.27, p=0.71). The mean methylation of DMR probes is also higher than other FDR significant probes which are not in DMRs (9.0%, $p=3.05\times10^{-5}$, see figure 3.10).

In keeping with the earlier observation that hypermethylated probes are more frequently found in CGIs, in 11/19 (57.9%) DMRs, the most significant probe

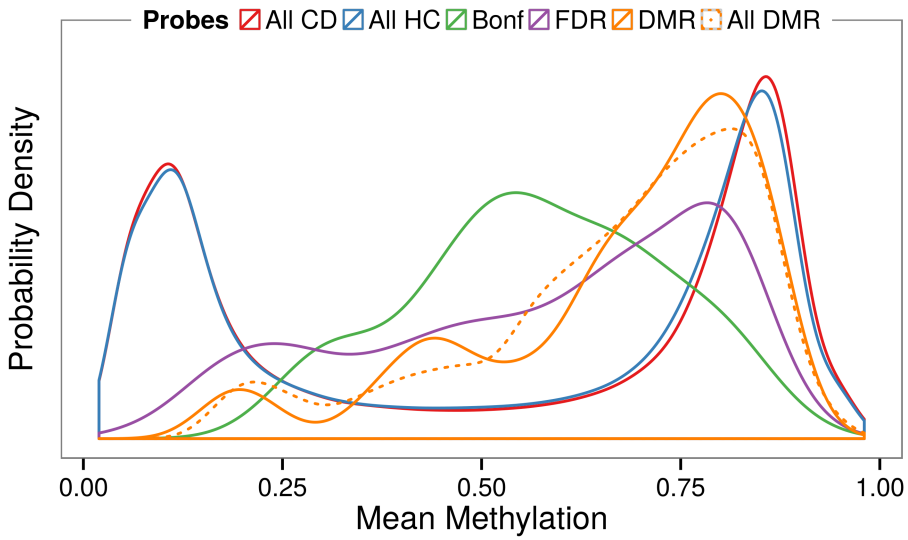


Figure 3.10: Mean methylation values for top (solid) and all (dotted) DMR probes, compared with all FDR significant probes, Bonferroni significant probes, and all probes separated by diagnosis

was within a CGI (see figure 3.11), compared to 19.4% (χ^2 p=2.8×10^{-4}) for non-DMR FDR significant probes, 31.1% (p=0.07) for Bonferroni significant probes and 60.7% (p=0.99) for all probes. 17/19 (89%) of DMRs intersect with a CGI.

Local context of DMRs

Figures 3.12–3.29 (pages 78–95) show the methylation results for each gene containing a DMR. From top: 1) -$\log_{10}$ p values, with lines indicating Bonferroni corrected p<0.05 and FDR<0.05. 2) Beta values for each sample, red - CD, blue - control. 3) Mean beta value for CD relative to mean control. 4) Gene transcripts. Grey highlight indicates the position of the DMR

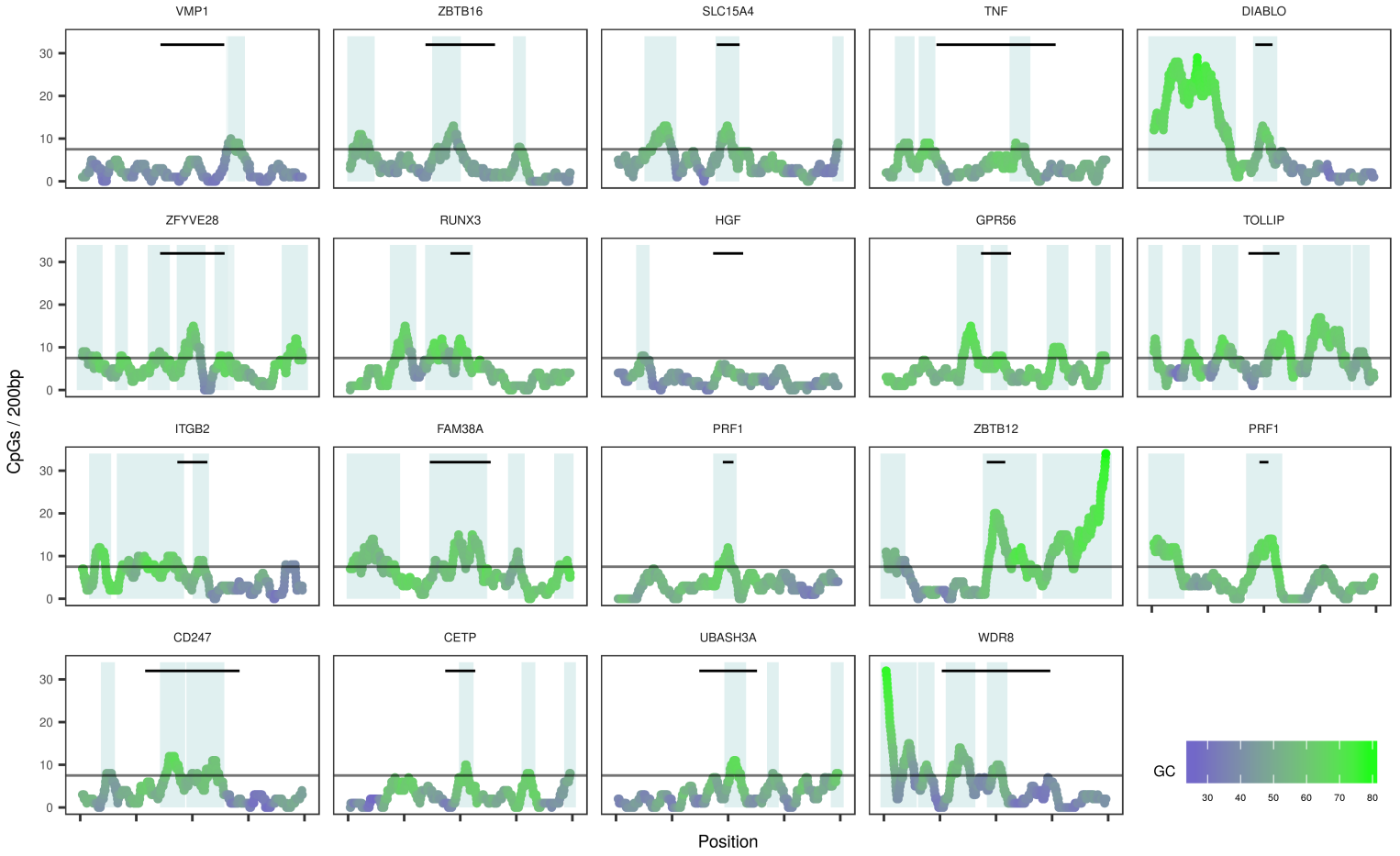


Figure 3.11: Number of CpGs and GC content per 200bp for DMR regions. Shaded areas represent CGIs, the lower horizontal line marks 60% of expected CpGs, x axis demarcations represent 1kb. The upper horizontal line represents the DMR boundaries derived from the lasso algorithm.

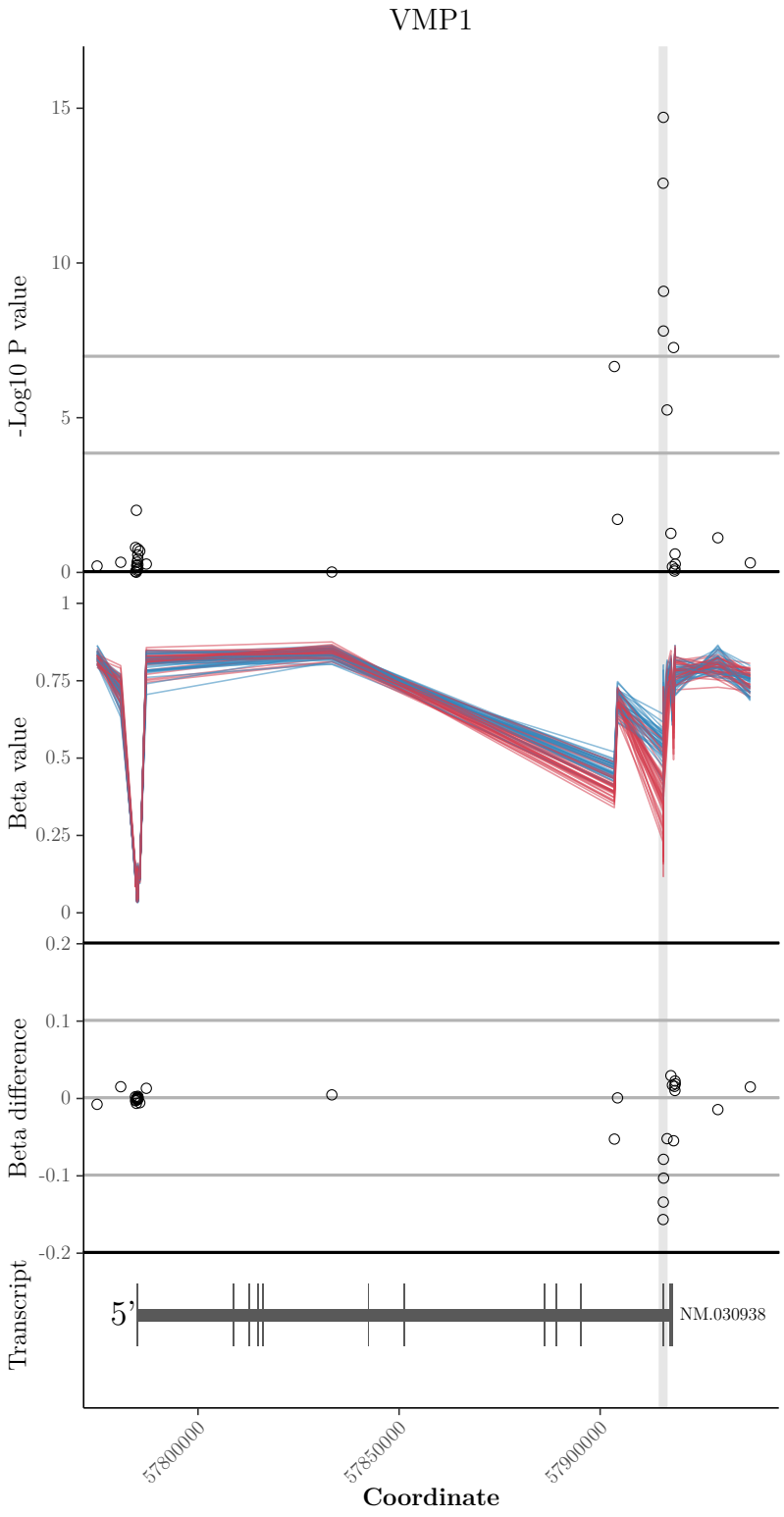


Figure 3.12: Methylation results for the VMP1 region. See page 76

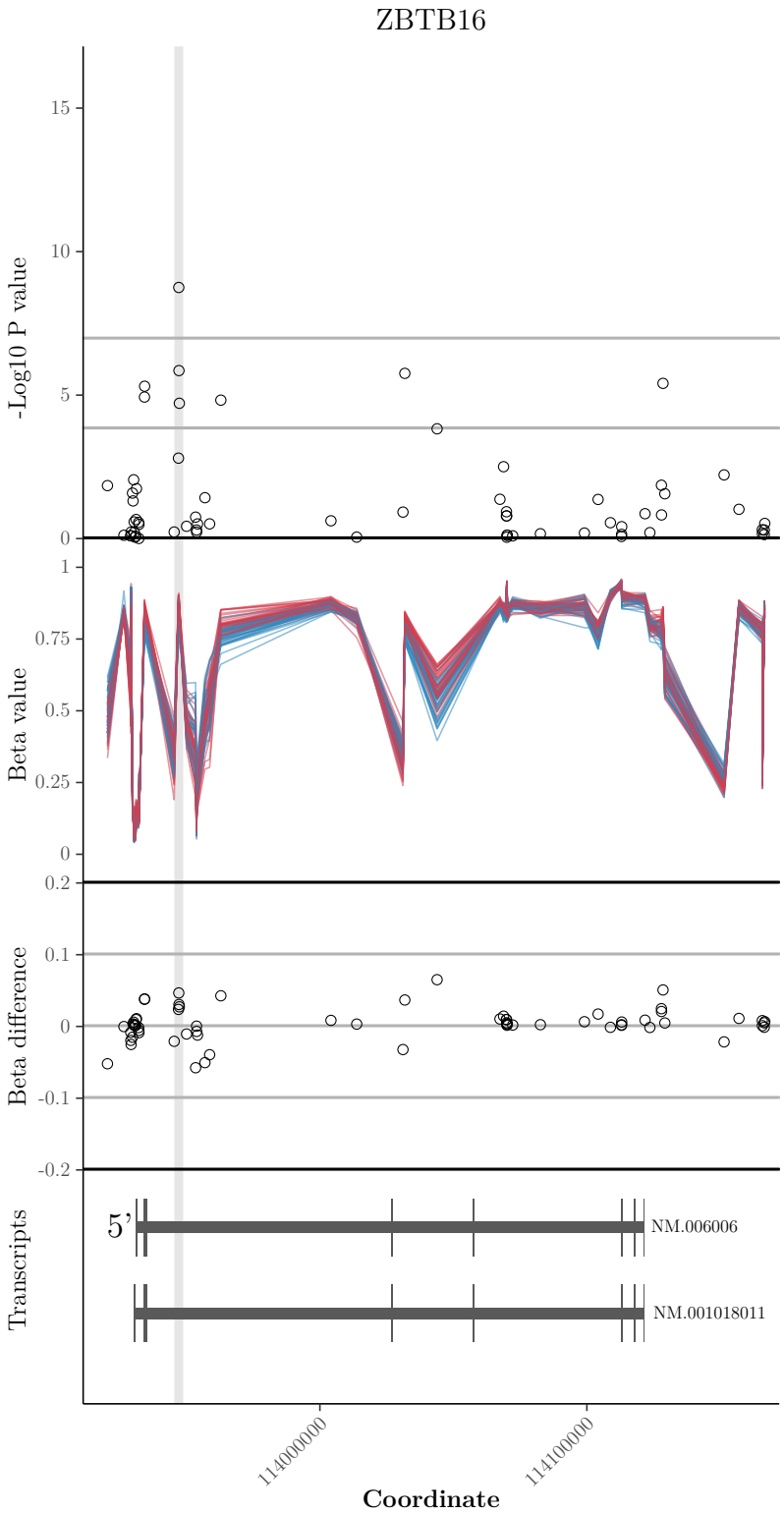


Figure 3.13: Methylation results for the ZBTB16 region. See page 76

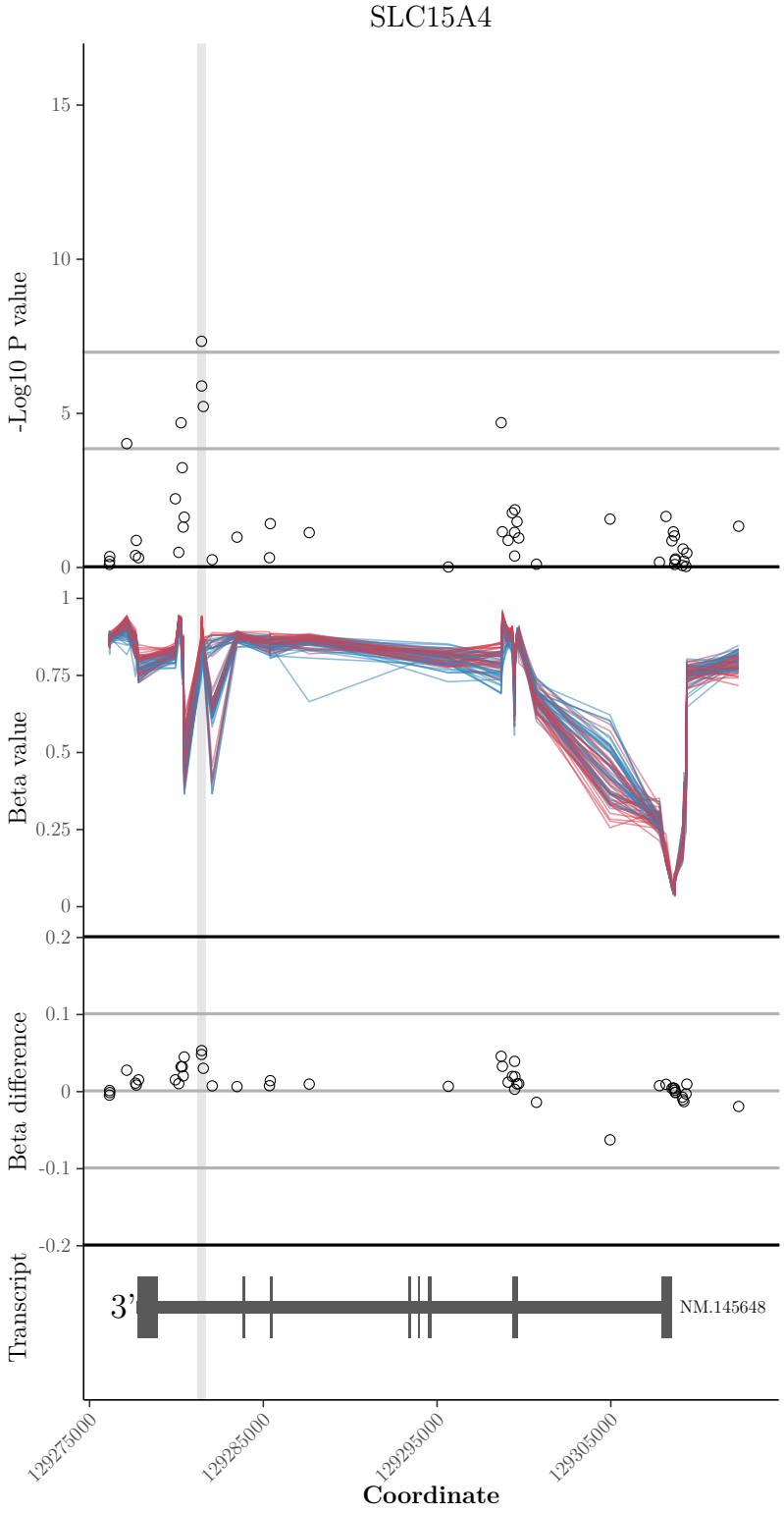


Figure 3.14: Methylation results for the SLC15A4 region. See page 76

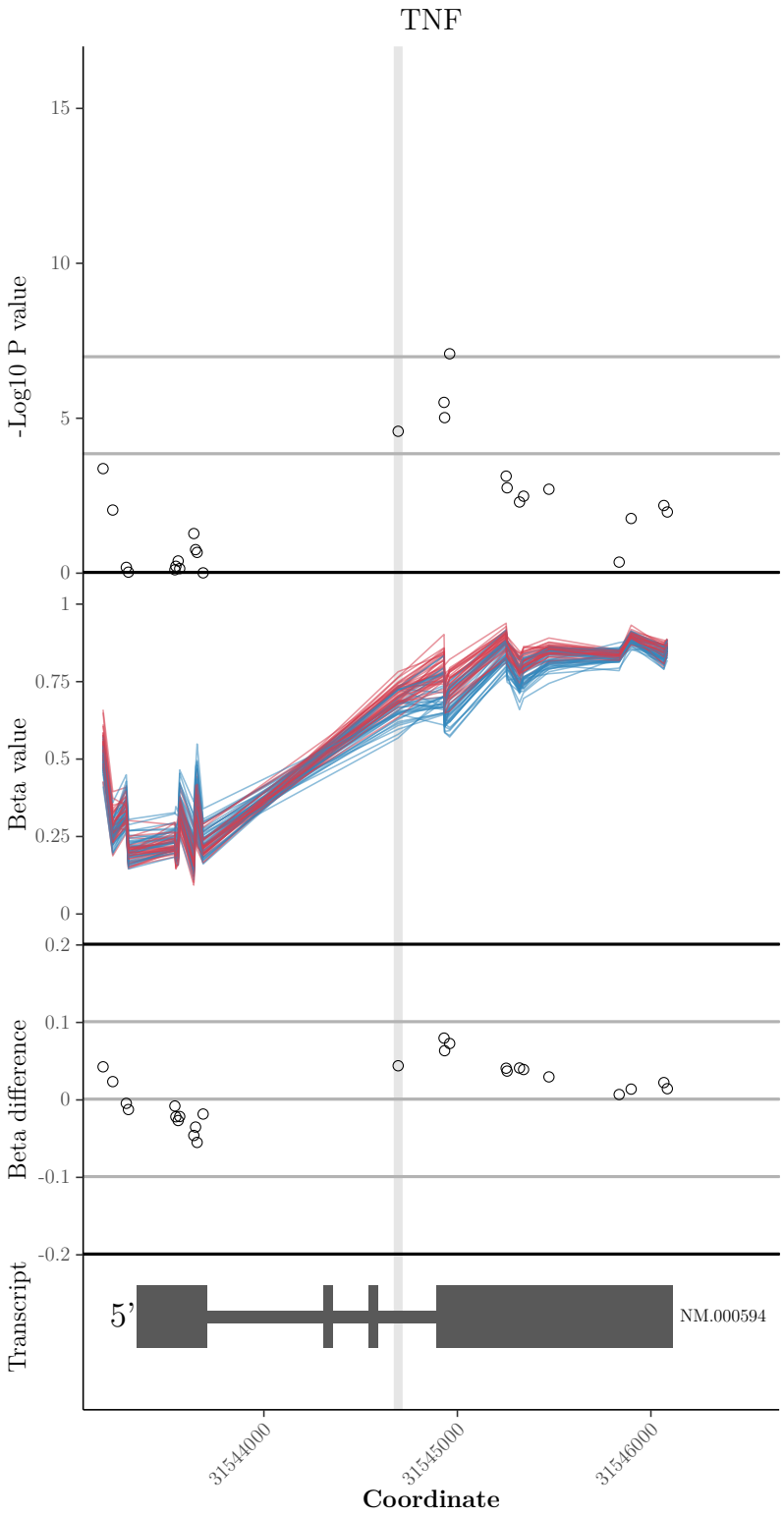


Figure 3.15: Methylation results for the TNF region. See page 76

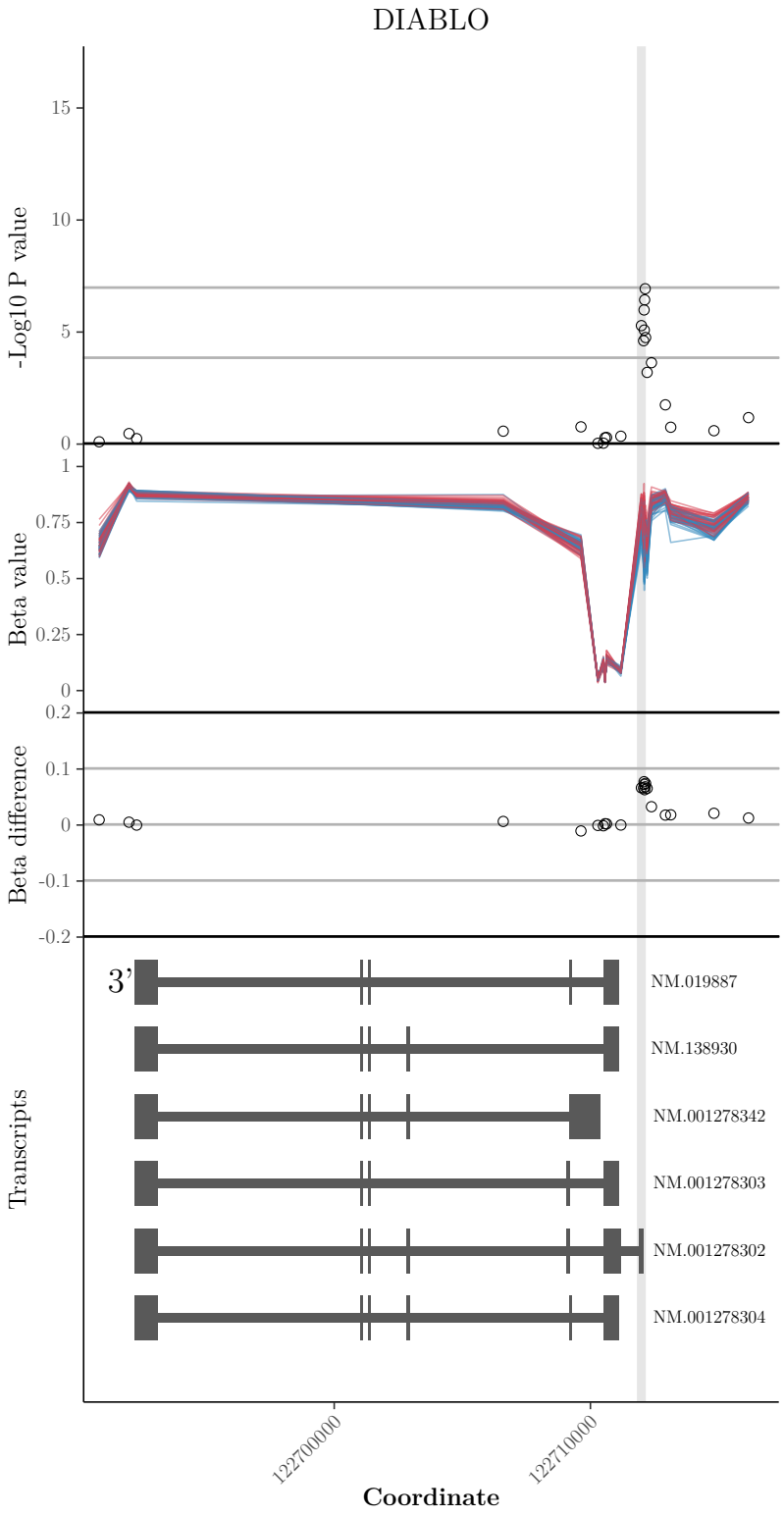


Figure 3.16: Methylation results for the DIABLO region. See page 76

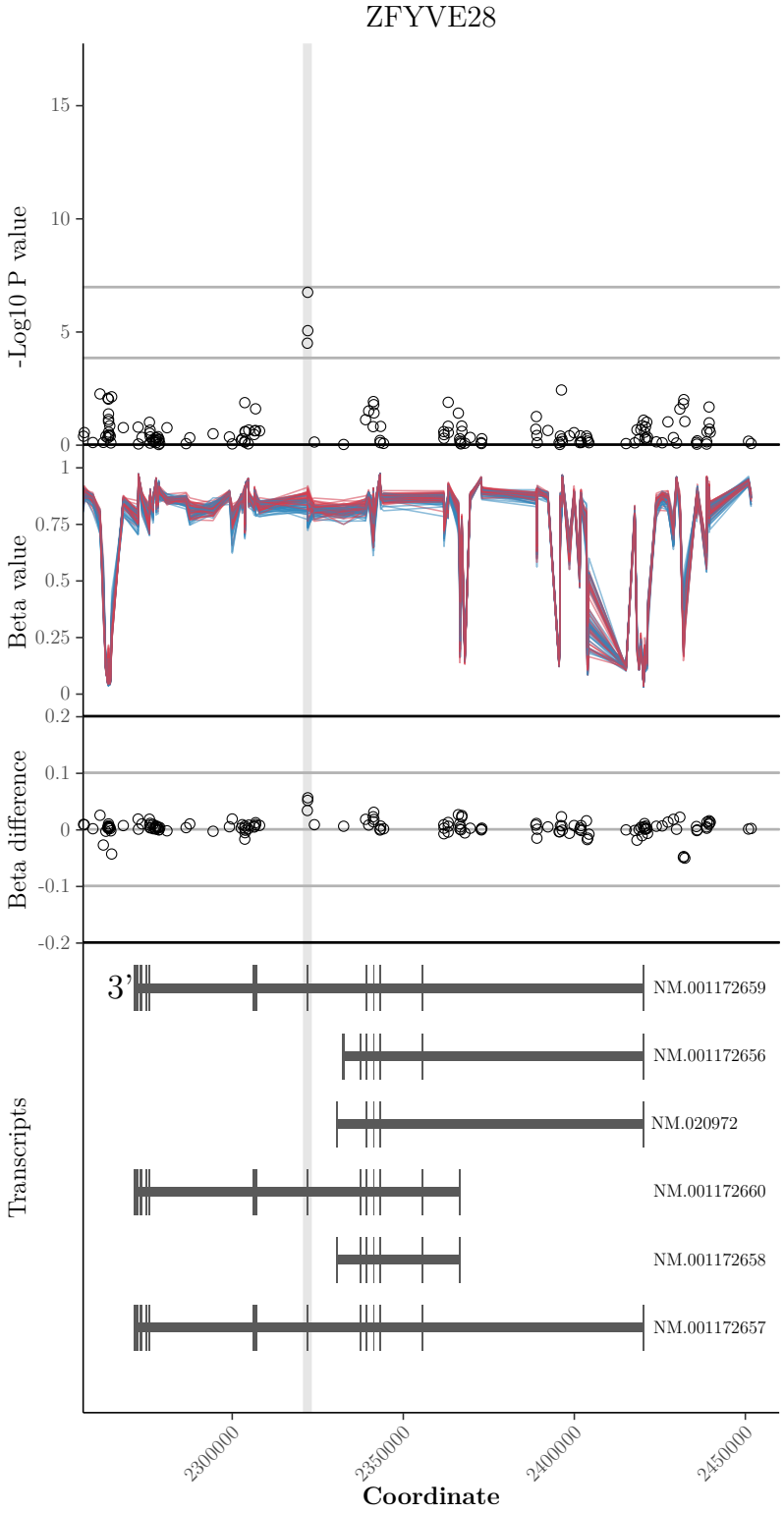


Figure 3.17: Methylation results for the ZFYVE28 region. See page 76

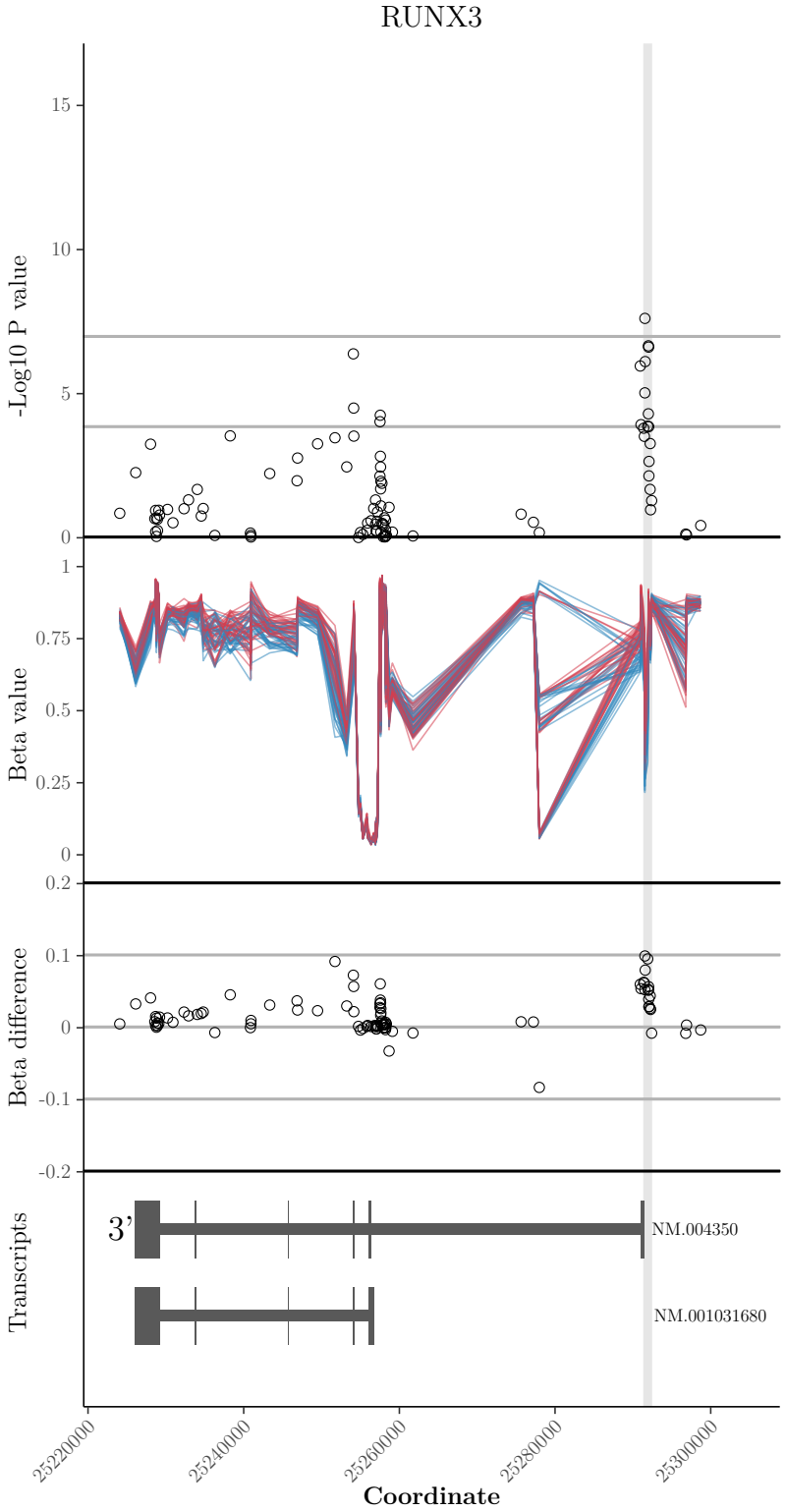


Figure 3.18: Methylation results for the RUNX3 region. See page 76

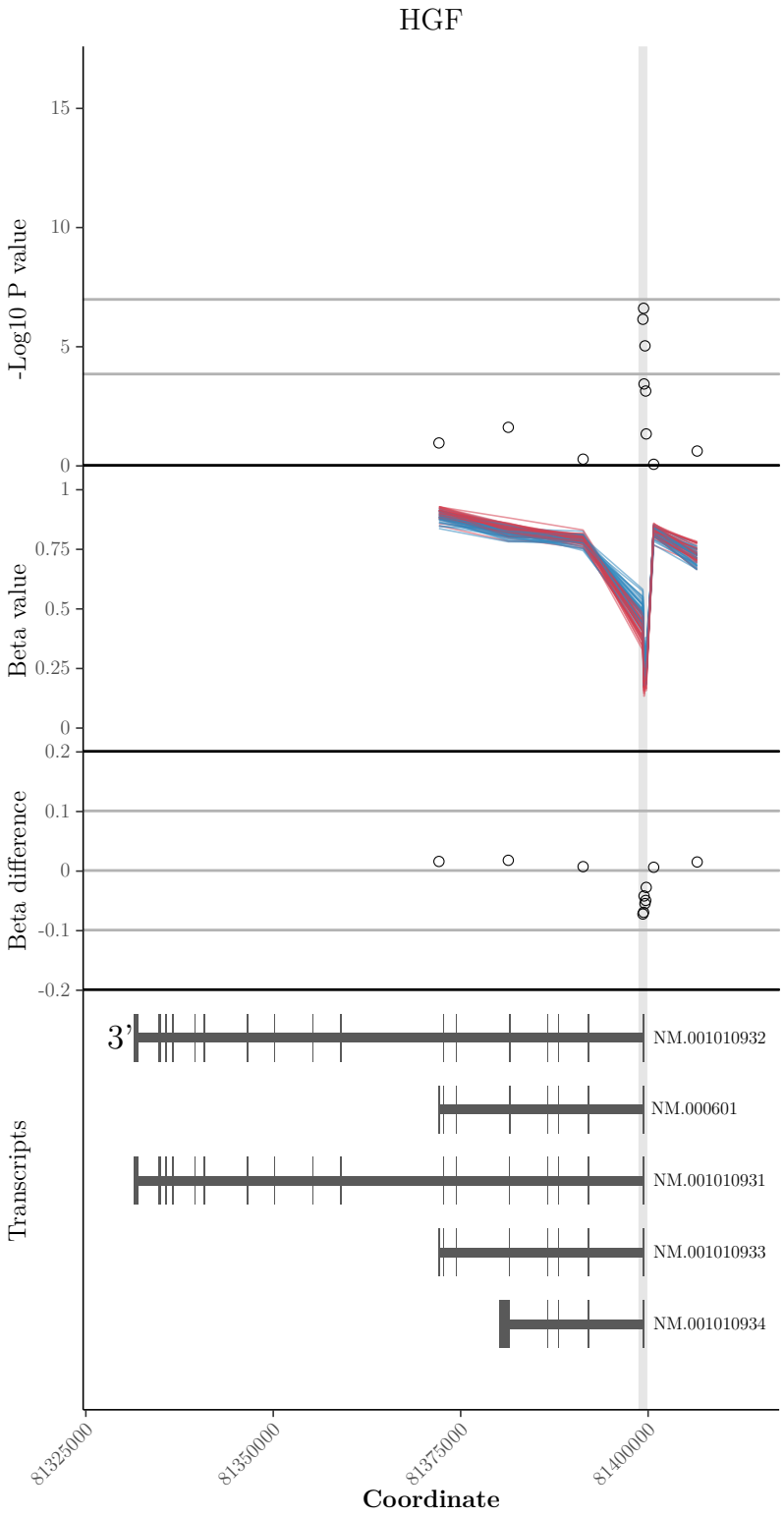


Figure 3.19: Methylation results for the HGF region. See page 76

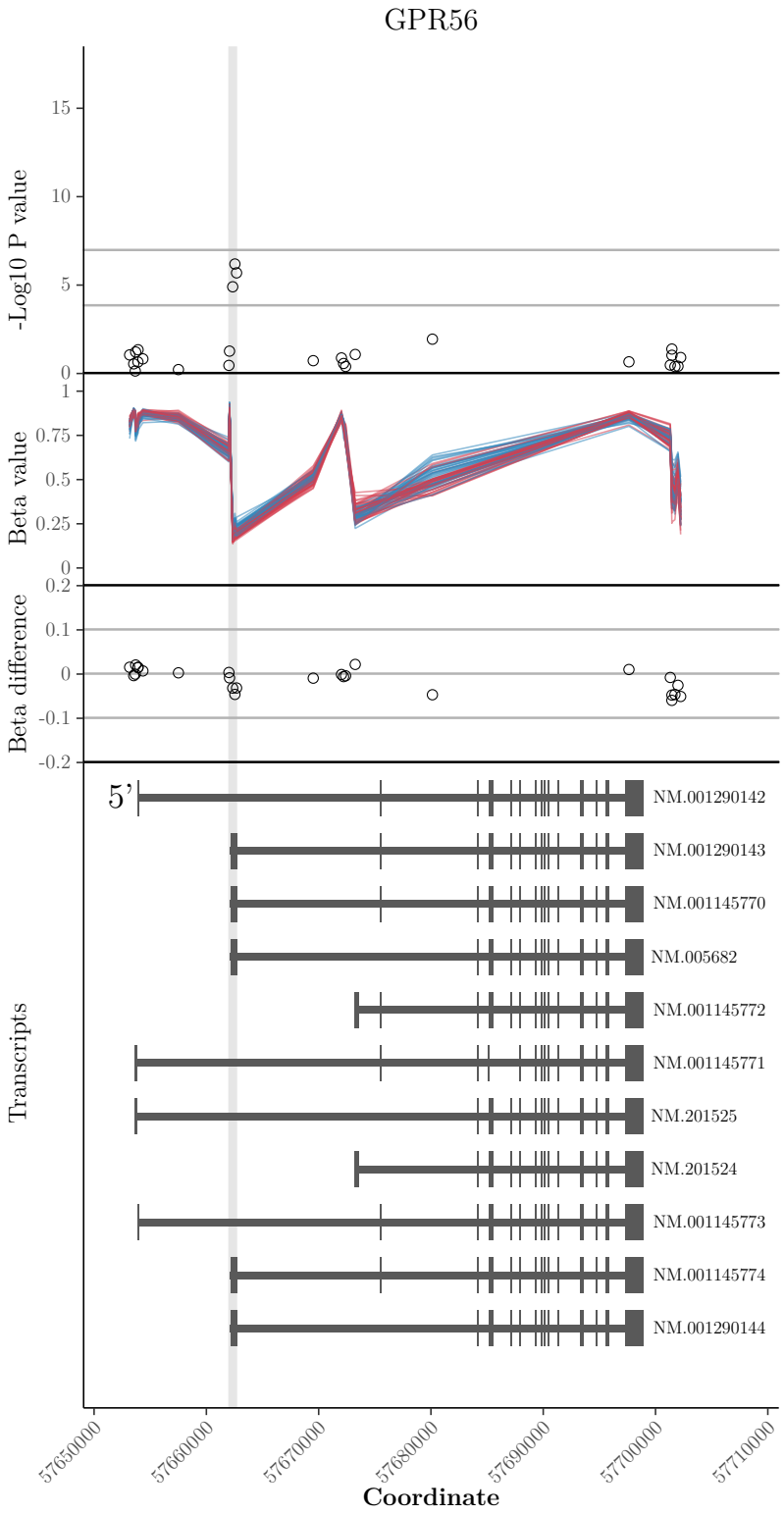


Figure 3.20: Methylation results for the GPR56 region. See page 76

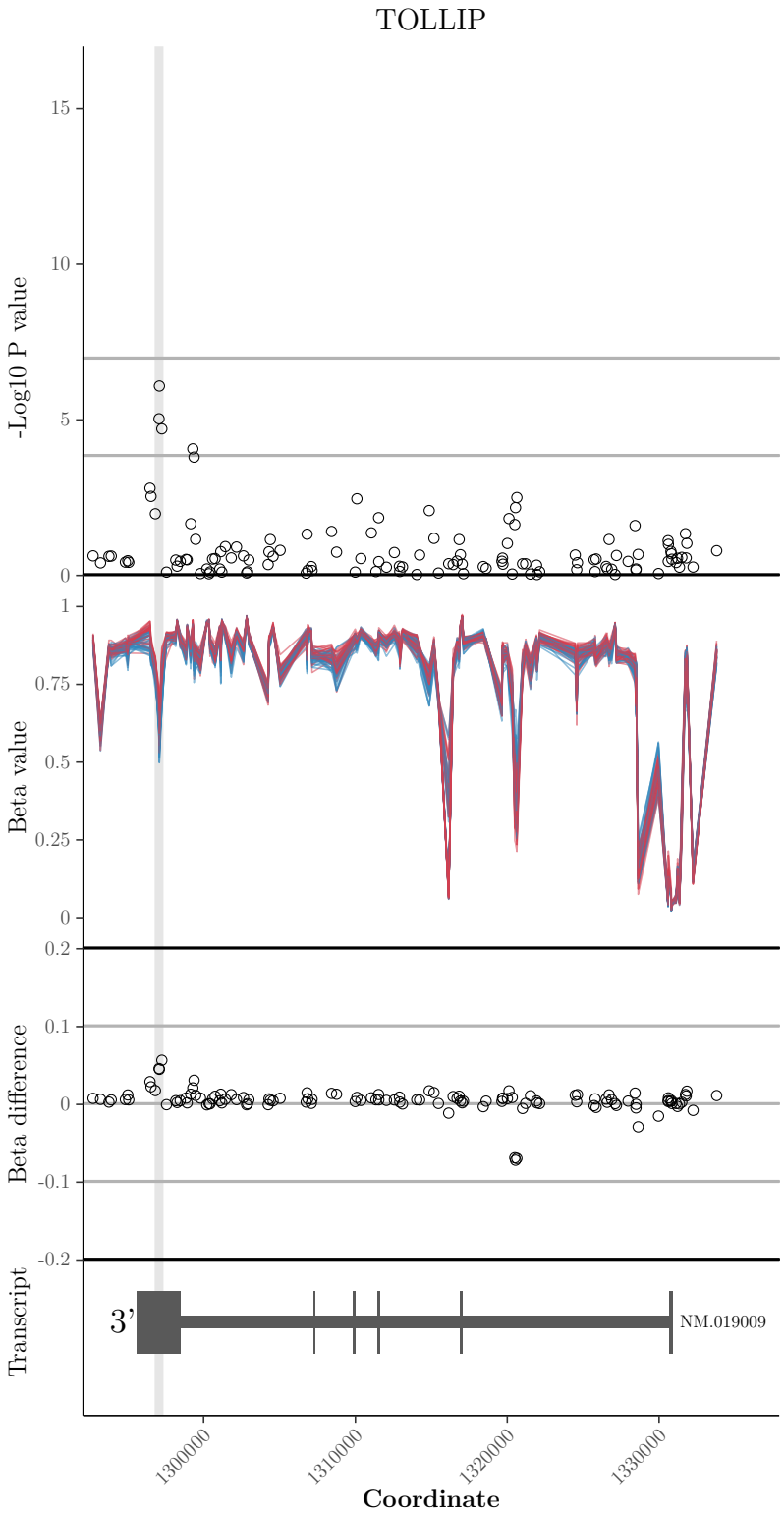


Figure 3.21: Methylation results for the TOLLIP region. See page 76

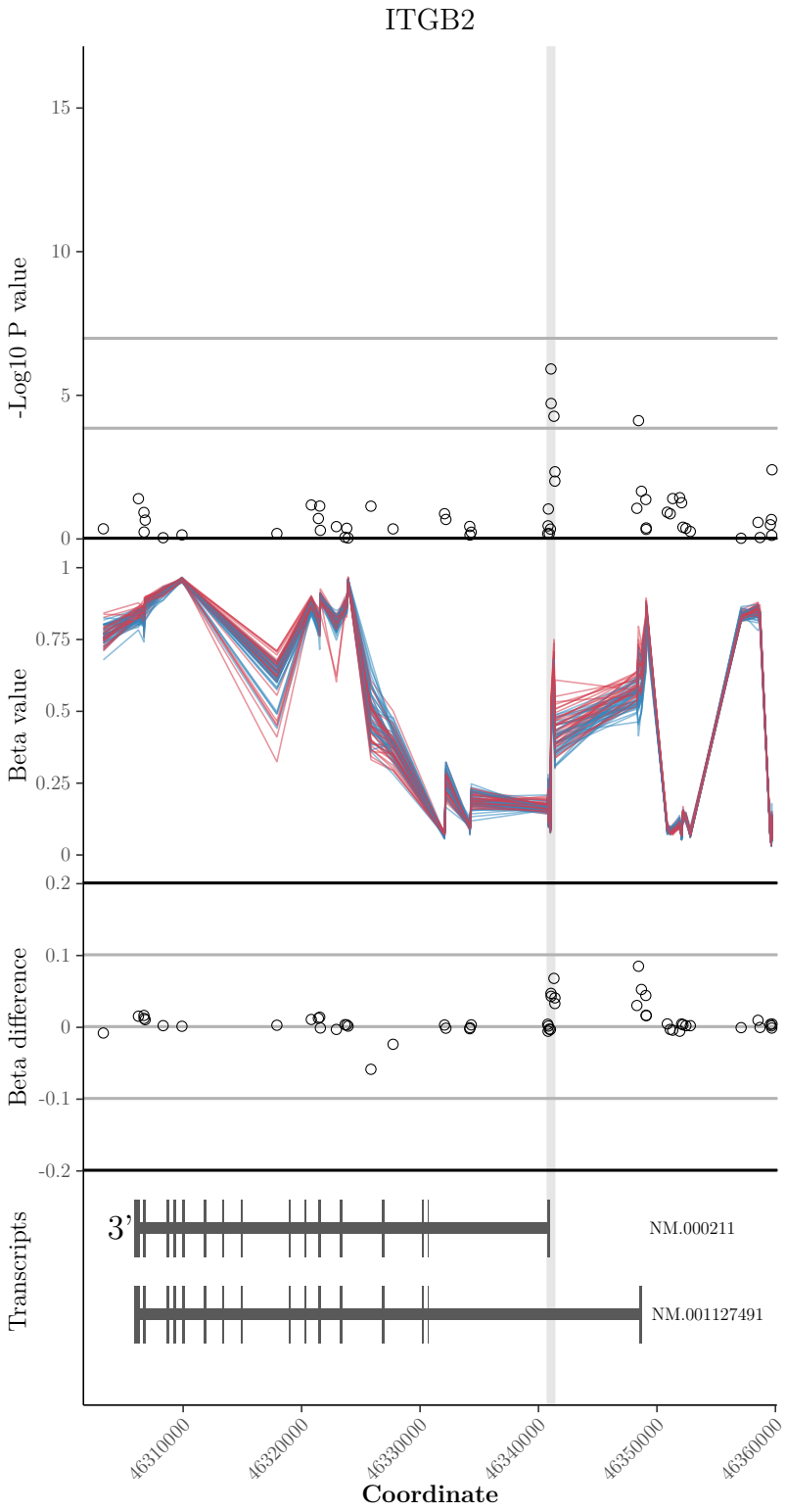


Figure 3.22: Methylation results for the ITGB2 region. See page 76

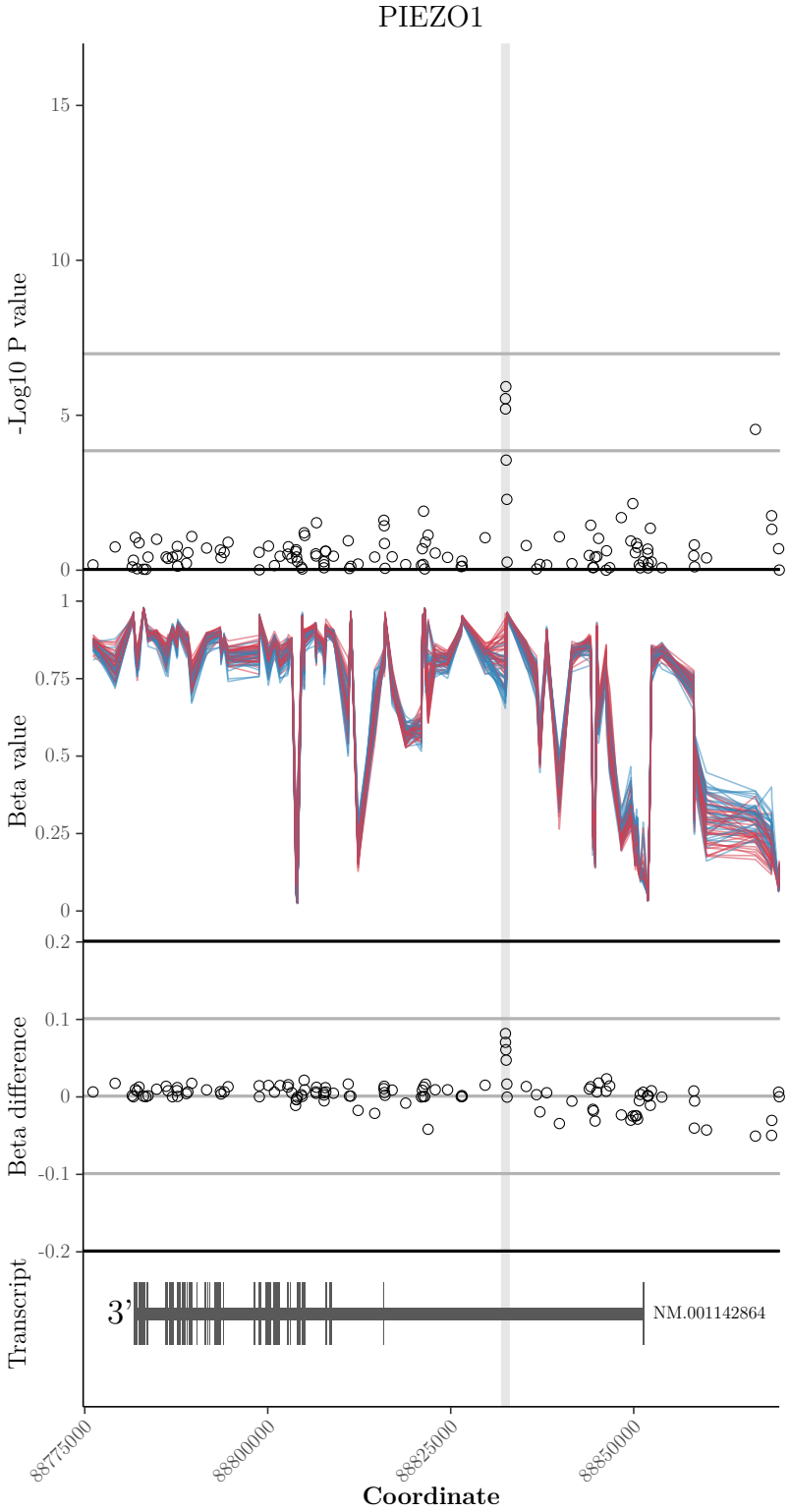


Figure 3.23: Methylation results for the PIEZO1 region. See page 76

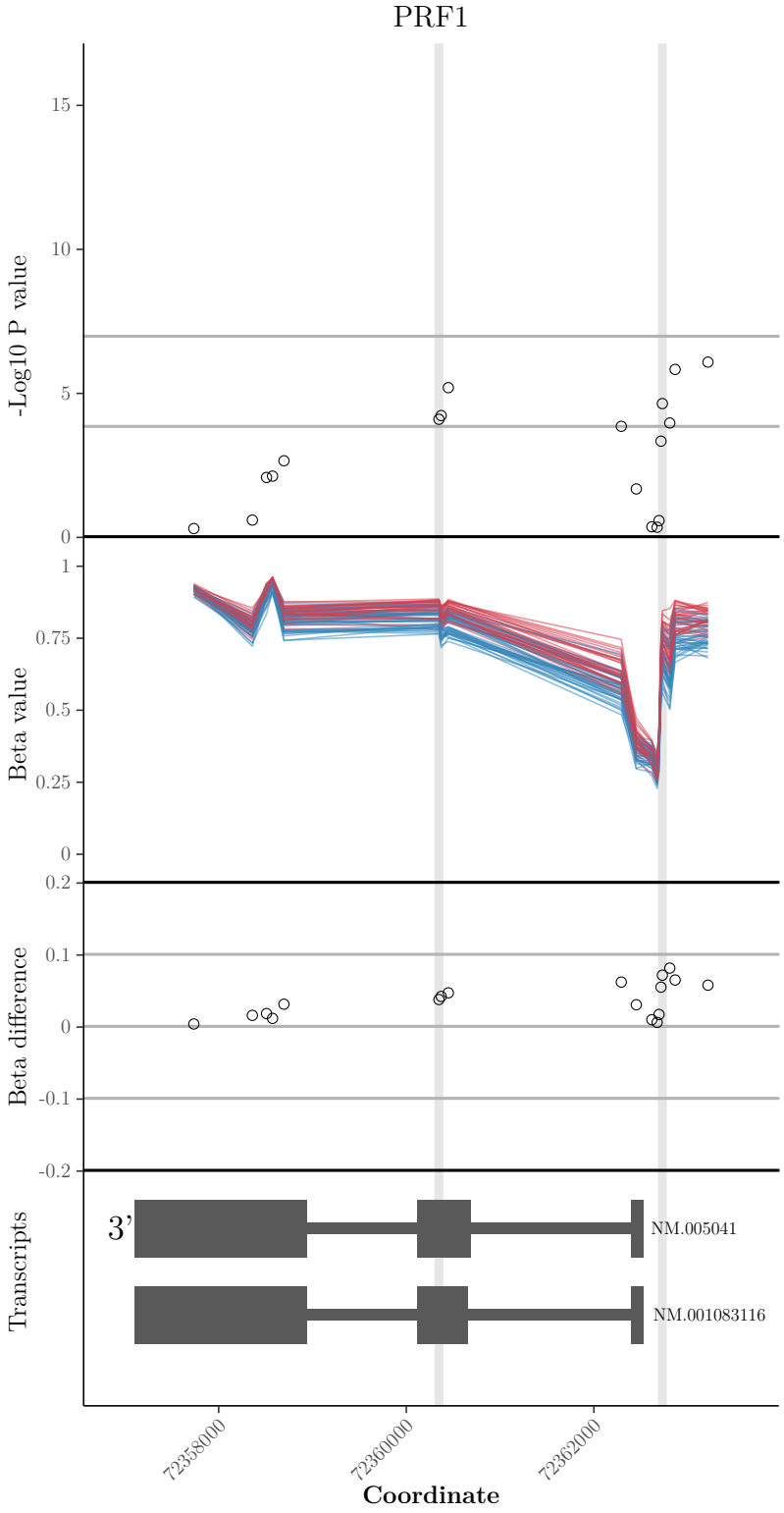


Figure 3.24: Methylation results for the PRF1 region. See page 76

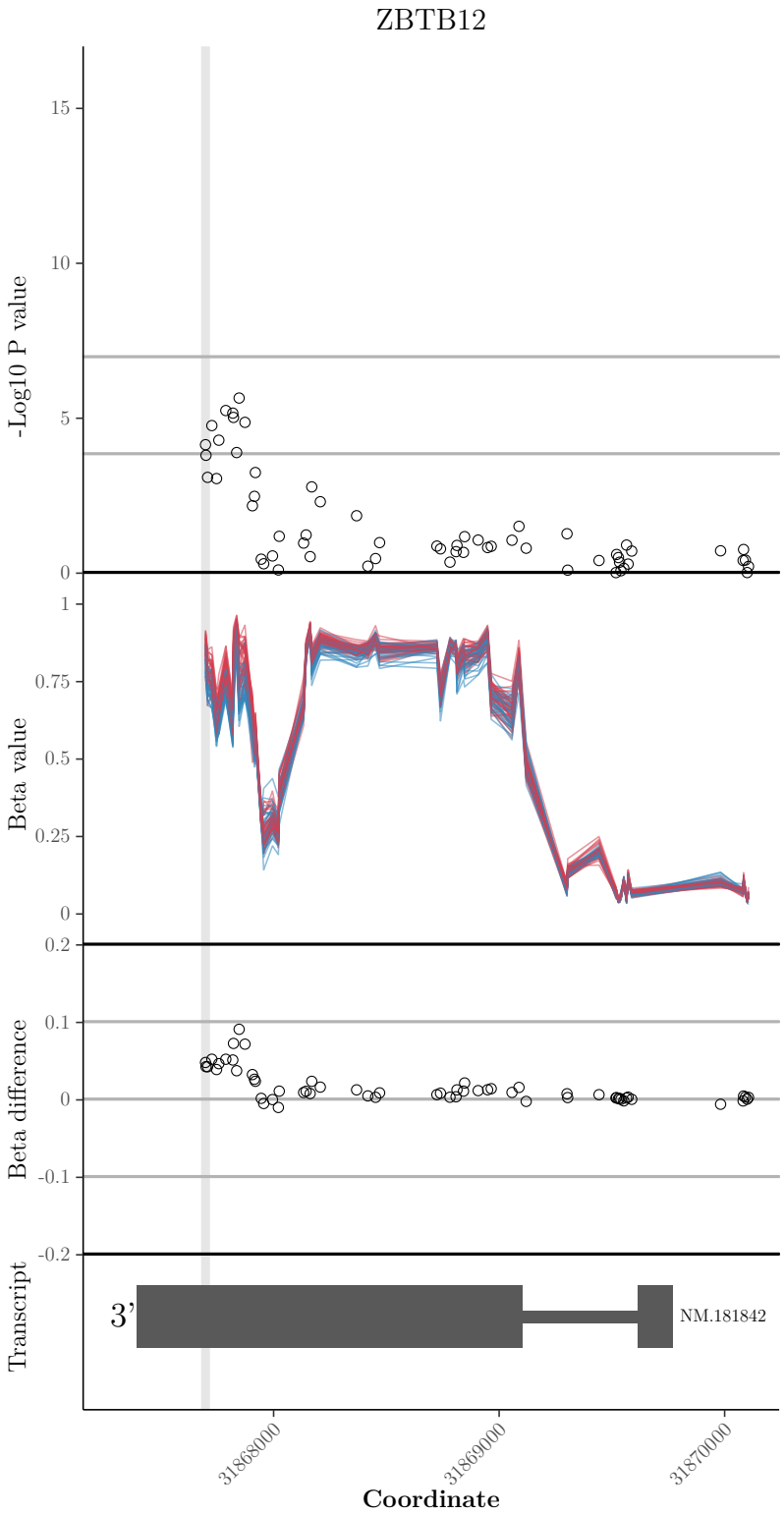


Figure 3.25: Methylation results for the ZBTB12 region. See page 76

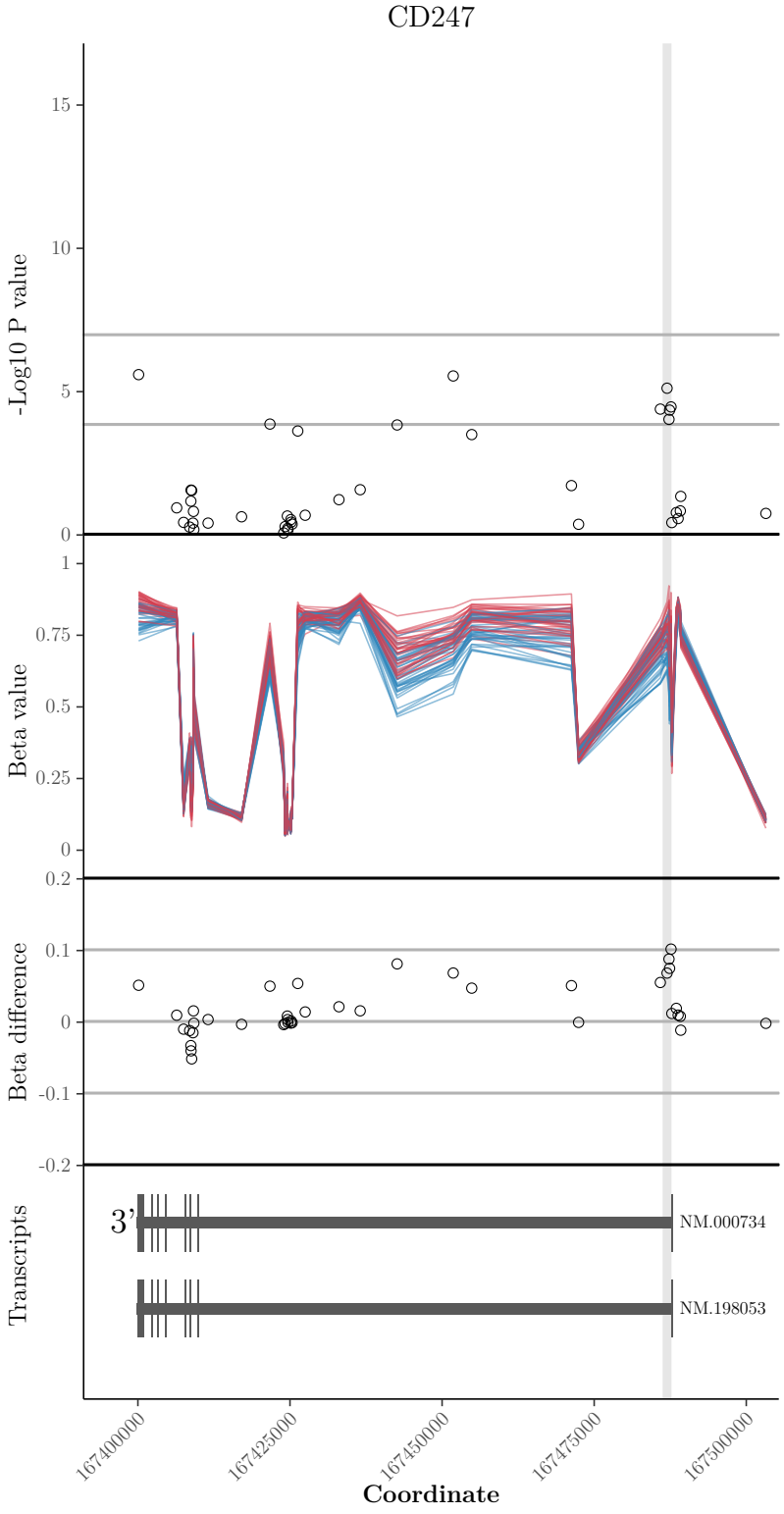


Figure 3.26: Methylation results for the CD247 region. See page 76

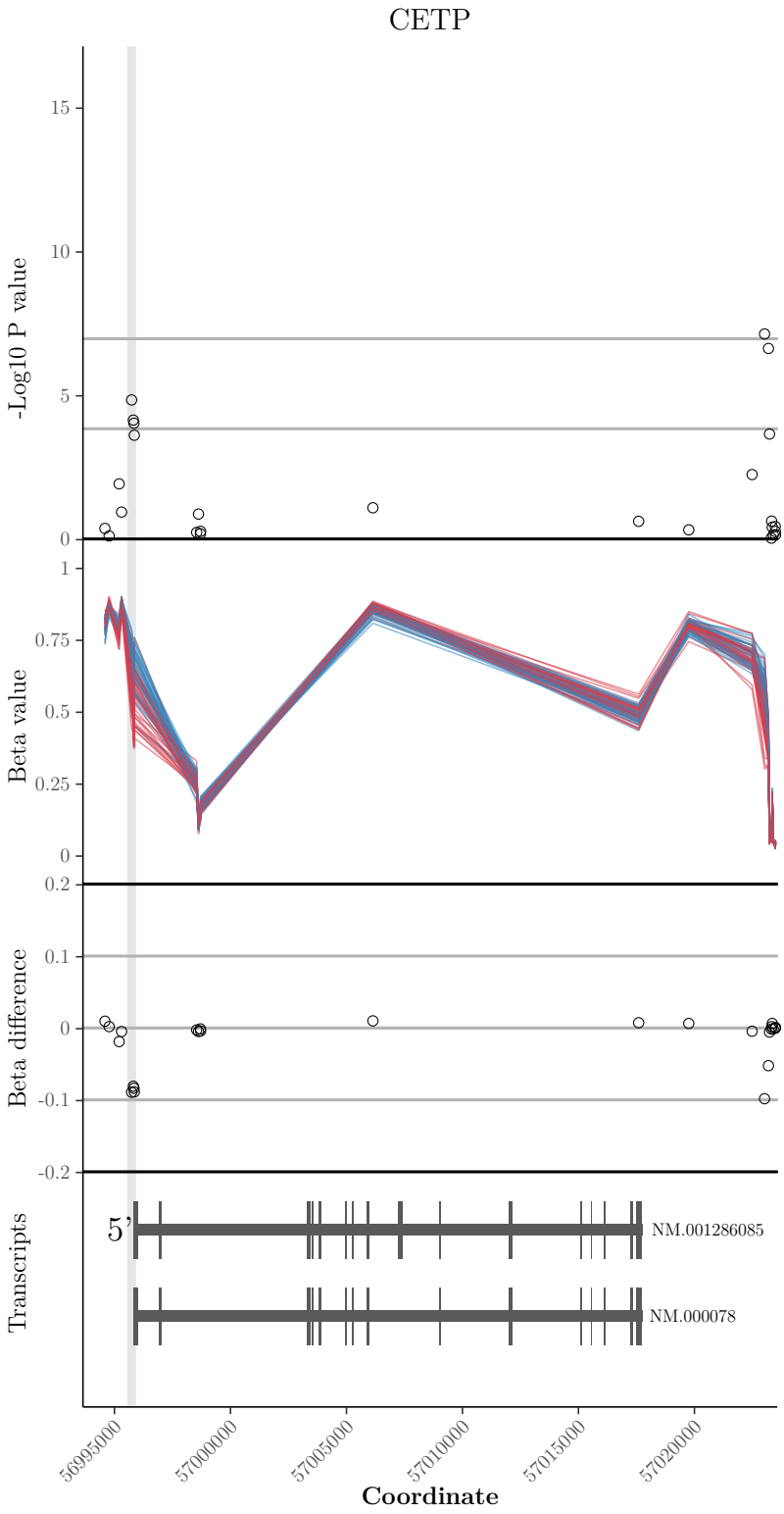


Figure 3.27: Methylation results for the CETP region. See page 76

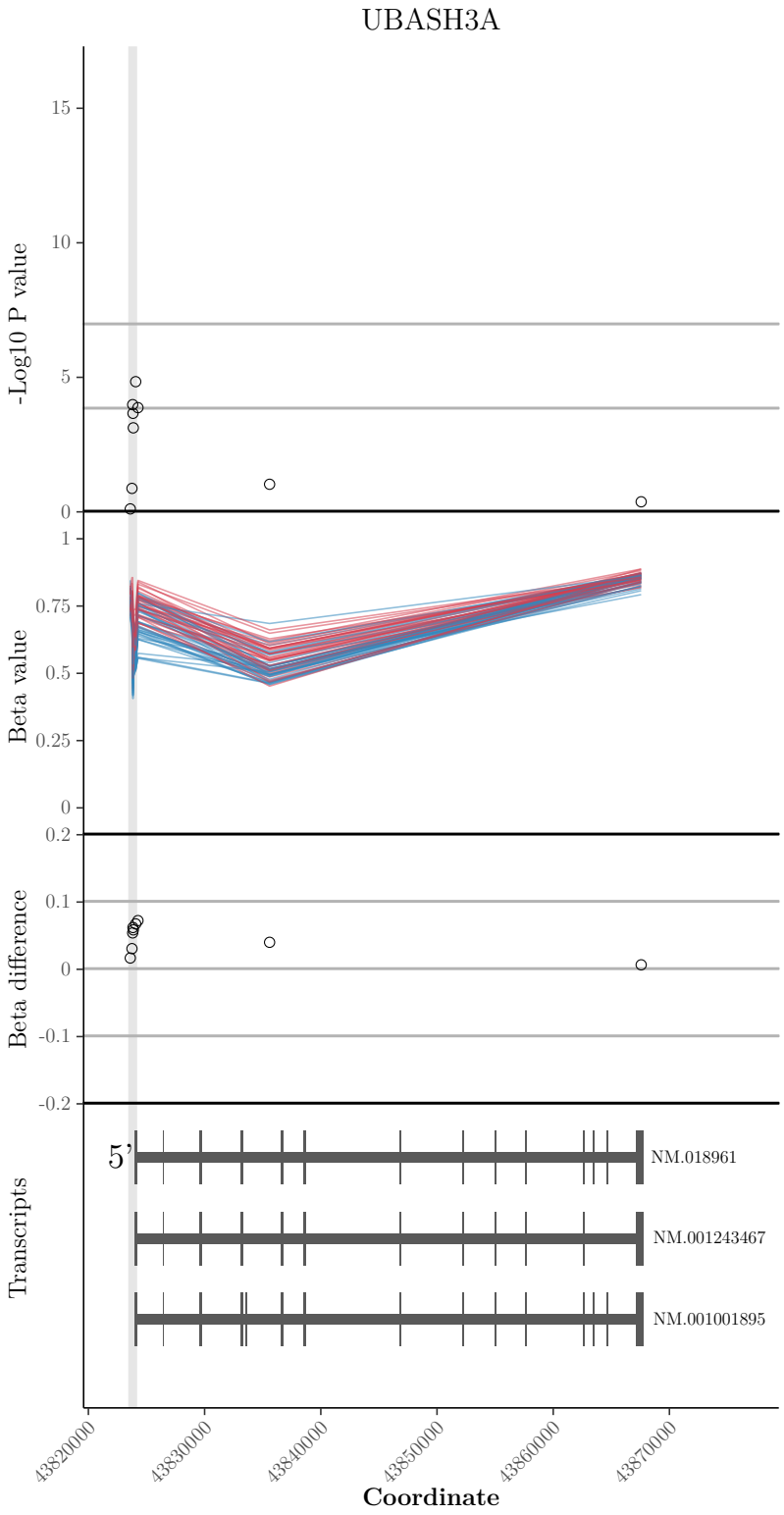


Figure 3.28: Methylation results for the UBASH3A region. See page 76

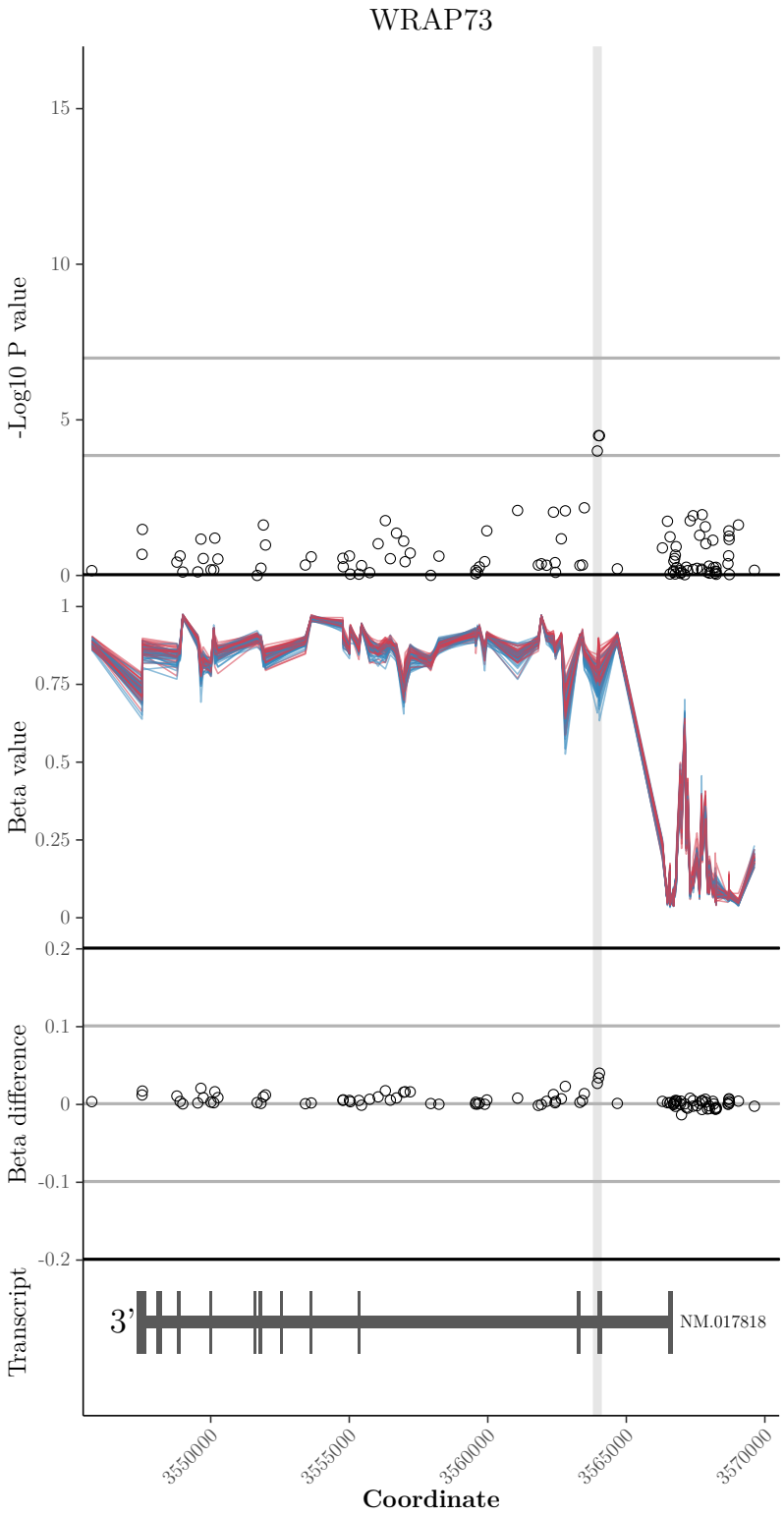


Figure 3.29: Methylation results for the WRAP73 region. See page 76

3.3.5 Correlation with GWAS

Analysis published in Adams *et al.* Using range thresholds of 25kb, 50kb, 100kb, 250kb, 500kb, 1mb, 2mb and 4mb, the p value of the most significant methylation probe within range of each GWAS SNP was recorded. The list of GWAS SNPs were drawn from Jostins *et al.* [297] and Liu *et al.* [384], excluding those which were solely associated with UC.

Across all distance thresholds there is a significant correlation between the number of probes within a bin and the probability of that bin containing a GWAS SNP. Likewise, bins with greater numbers of probes are more likely to contain more significant methylation probes. Therefore, for each range threshold 1000 random bins were selected which matched the distribution of Illumina 450k probe coverage seen in the CD/IBD SNPs. Matching was based on splitting the range of the square root of probes per bin into ten groups and establishing the proportions of bins falling into each group. Non-SNP bins were then chosen at random from groups of equivalent square root probe number with the same proportion per decile. Examples of the matching results are shown in figures 3.30 and 3.31 (page 97). The smallest p values within the randomly selected bins were then compared to those within range of CD/IBD SNPs by Wilcoxon rank-sum test.

Methylation results were more significant within proximity of CD/IBD GWAS SNPs in all tested range thresholds. The distance thresholds where the correlation was strongest were 50kb ($p = 3.66 \times 10^{-7}$) and 100kb ($p = 2.41 \times 10^{-7}$). This result was presented in Adams *et al.* [4], with a figure (see figure 3.32) showing a similar analysis to that of Nimmo *et al.* [458] — *i.e.* for methylation results with lower p values there is an increasing likelihood of being near a GWAS SNP.

Illumina 450k probes with significant methylation changes after Bonferroni correction which are within 1Mb of IBD GWAS SNPs from Jostins *et al.* and Liu *et al.* are shown in table 3.11 on page 99. There were 6 DMRs (table 3.10, page 75) within 1Mb of these SNPs; *VMP1*, *ZBTB16*, and *TNF* are already shown with other Bonferroni significant probes in table 3.11, additionally *TOLLIP*, *ITGB2*, and *ZBTB12* were 541, 690, and 593kb from an IBD GWAS SNP respectively.

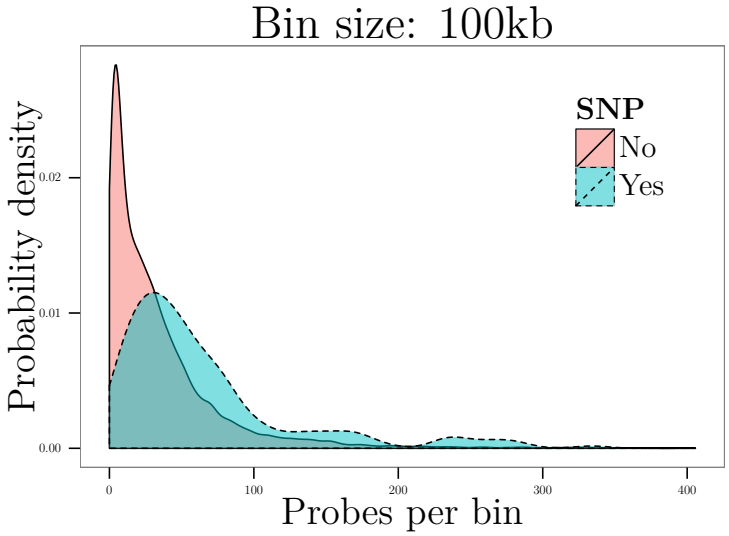


Figure 3.30: Probe Density is higher in proximity to IBD GWAS SNPs compared to unmatched regions

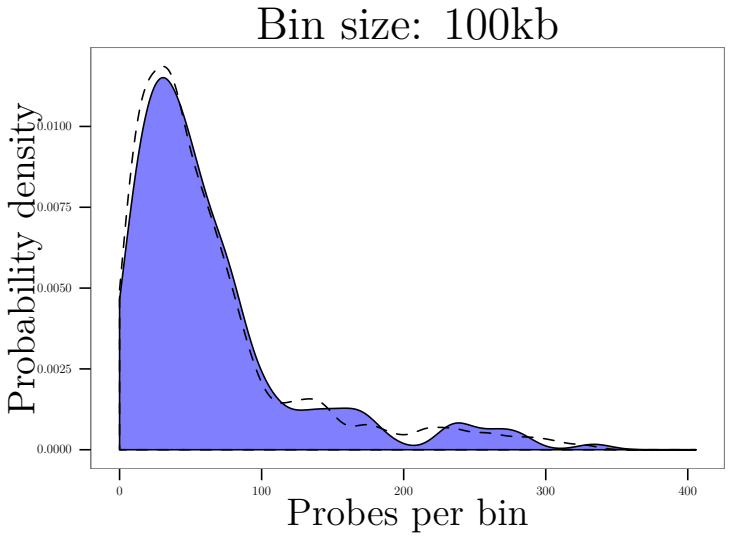


Figure 3.31: Results of Probe Density Matching. Density of probes within 100kb of IBD GWAS SNPs shown in blue, with the dotted line showing a matched set of random probes

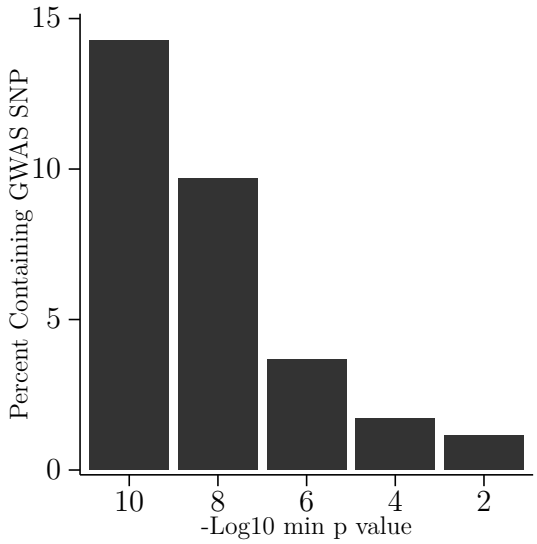


Figure 3.32: Enrichment of methylation changes within 50kb of GWAS SNPs (from Adams *et al.* [4])

Gene	Chr:Coord	P value[a]	GWAS SNP	Distance	Probes[b]	Candidate genes[c]
SBNO2	19:1130866	2.26×10^{-3}	rs2024092	6835	2	GPX4,HMHA1,(20)
IL18RAP	2:103038171	3.00×10^{-3}	rs917997	32397	1	IL18RAP,IL1R1,(7)
VMP1	17:57915717	8.86×10^{-10}	rs1292053	47820	5	TUBD1,RPS6KB1,(9)
RTP5	2:242813914	5.51×10^{-3}	rs35320439	76573	1	PDCD1,ATG4B
NA	1:151945663	1.14×10^{-4}	rs4845604	143983	1	RORC
GSDMC	8:130799007	1.42×10^{-3}	rs1991866	174902	1	(2)
CLU	8:27467783	2.12×10^{-3}	rs17057051	240229	1	PTK2B,TRIM35,EPHX2
TNF	6:31544960	3.73×10^{-2}	rs9264942	270580	1	(22)
PTGER4	5:40685989	8.45×10^{-3}	rs11742570	275405	1	PTGER4,(1)
NRXN2	11:64428925	2.40×10^{-3}	rs559928	278555	1	CCDC88B,(23)
ZEB2	2:145172035	1.16×10^{-3}	rs11681525	320347	1	-
NA	15:41233701	1.60×10^{-3}	rs28374715	330249	1	(11)
NA	3:101901234	7.13×10^{-8}	rs616597	331508	1	NFKBIZ
SOCS3	17:76354621	2.04×10^{-4}	rs17736589	382497	1	-
RPS6KA2	6:166970252	2.01×10^{-9}	rs1819333	403295	1	CCR6,RPS6KA2,(4)
ZBTB16	11:113947148	8.01×10^{-4}	rs561722	439682	1	NXPE1,NXPE4,(5)
ZC3H4	19:47618017	3.40×10^{-2}	rs1126510	494234	1	CALM3,(14)
NPDC1	9:139939792	4.05×10^{-2}	rs10781499	673387	1	CARD9,(22)
CBFA2T2	20:32235621	2.91×10^{-2}	rs4911259	859339	1	DNMT3B,(8)

Table 3.11: Bonferroni Significant Probes within 1Mb of GWAS SNPs from Jostins *et al.* [297] and Liu *et al.* [384]. [a] Bonferroni corrected [b] Bonferroni significant probes within the same gene. [c] Candidate genes identified by the reporting papers, (additional genes), - no reported candidates

Re-analysis of correlation Although the randomly selected bins were matched for probe density, there may be other factors making the regions surrounding GWAS SNPs different from the randomly selected regions. To attempt to remove these uncontrolled variables the significance of methylation findings in proximity to IBD SNPs was compared to the methylation significance in proximity to randomly selected GWAS SNPs for other conditions using the NHGRI-EBI Catalog of published genome-wide association studies to source both the IBD and non-IBD SNPs [264].

1000 randomly selected SNPs were chosen to match the density of Illumina 450k coverage and the minimum p value for CD-associated methylation differences within a range of distances was compared to those seen in IBD GWAS SNPs. This analysis was performed on 1000 random selections of non-IBD GWAS SNPs. The summary results of the 1000 repeats are shown in figure 3.33 (page 101), median p values for enrichment of methylation results in proximity to IBD GWAS SNPs at 50kb at 100kb bin sizes are 5.89×10^{-5} and 1.19×10^{-5}.

To demonstrate that the randomly selected regions from the initial analysis are an inadequate control, the minimum methylation p value in 1000 random regions was compared with the minimum methylation p value in range of all non-IBD GWAS SNPs from the NHGRI-EBI Catalog. This showed an enrichment of significant methylation findings, particularly at 25–50kb (figure 3.34, page 102).

GWAS SNPs for inflammatory and autoimmune traits and some other traits with extensive GWAS results were also analysed. For each trait the most significant methylation results within the GWAS SNP bins were compared to 1000 randomly selected bins containing GWAS SNPs for other non-IBD traits, with the probe density for the two groups matched. This was repeated for 1000 random selections of control SNPs. The results are shown in figures 3.35 and 3.36 (pages 104 and 105).

There was a strong correlation between the methylation results and GWAS SNPs for HDL cholesterol over a wide range of distance thresholds, with weaker correlations for LDL cholesterol and lipid metabolism phenotype. Of the inflammatory and immune diseases the strongest correlation was for rheumatoid arthritis, with correlations also seen in type 1 and 2 diabetes, multiple sclerosis and systemic lupus erythematosus.

There was a weak but significant positive correlation between the P value of GWAS SNPs for all conditions and the significance of nearby methylation findings across all bin sizes (see table 3.12).

When comparing the p values of GWAS SNPs for individual traits only

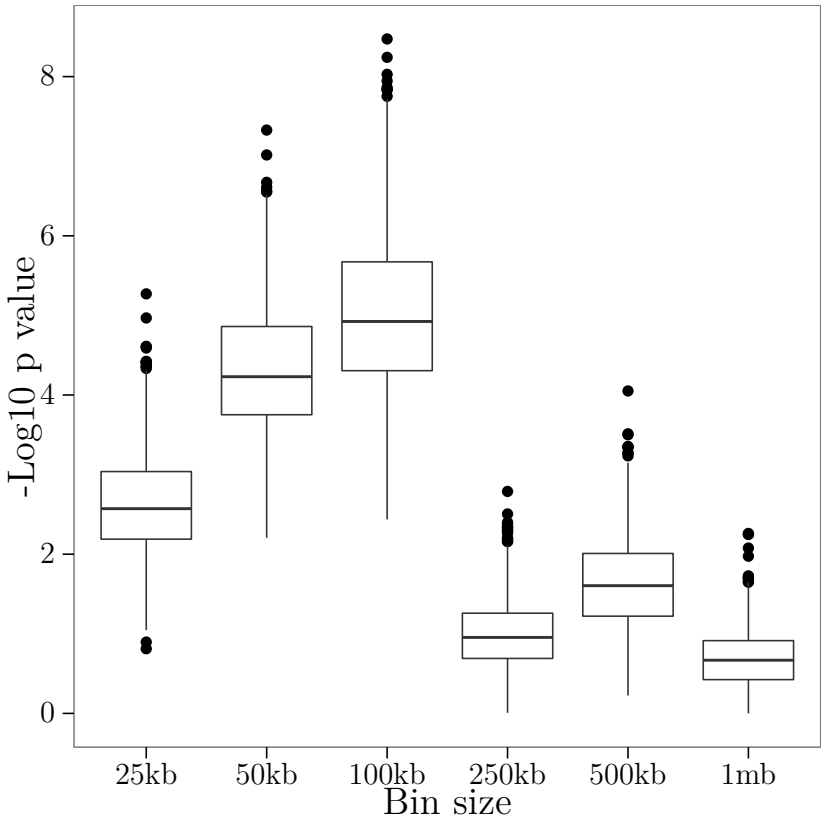


Figure 3.33: Box and whisker plot showing that CD-associated methylation differences are more likely to be found in proximity to IBD gwas SNPs than randomly selected GWAS SNPs for other traits, with a peak at distances around 100kb. Whiskers extend up to 1.5 $\times$ IQR from the 1st and 3rd quartiles, with any remaining results plotted as points. Summary results of 1000 random selections of non-IBD GWAS SNPs from the NHGRI-EBI GWAS Catalog matching the Illumina 450k coverage of IBD GWAS SNPs.

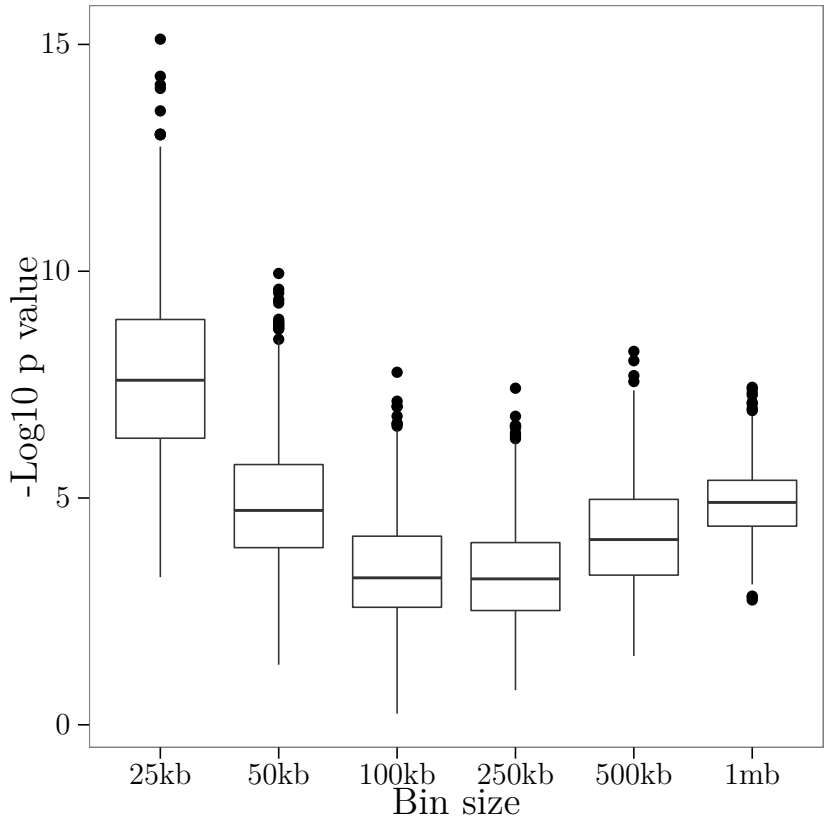


Figure 3.34: Box and whisker plot showing that significant methylation findings are enriched in proximity to all non-IBD GWAS SNPs compared to randomly selected regions of matching probe density. Whiskers extend up to 1.5 $\times$ IQR from the 1st and 3rd quartiles, with any remaining results plotted as points.

Bin size	p value	ρ
25kb	2.90×10^{-167}	0.21
50kb	2.33×10^{-181}	0.22
100kb	1.12×10^{-168}	0.21
250kb	4.62×10^{-206}	0.23
500kb	1.67×10^{-165}	0.21
1Mb	1.69×10^{-118}	0.18

Table 3.12: More significant GWAS SNPs for all diseases collectively, have more significant methylation findings in proximity to them. ρ: Spearman's rank correlation rho

two showed a significant correlation[1] with methylation p values; IgG glycosylation (p=0.014, Spearman's ρ=0.16, 699 SNPs) and lipid metabolism phenotypes (p=0.0036, ρ=0.46, 91 SNPs).

Correlation with predicted functional SNPs Farh *et al.* [176] attempted to predict functional variants in proximity to IBD GWAS SNPs. There was no enrichment of significant methylation results in proximity to these SNPs at any tested bin range by χ^2 test for Bonferroni or FDR significance (p=1) and no significant correlation between methylation p value and distance to the nearest predicted functional variant by Spearman's rank correlation (p=0.34–0.95).

[1]Bonferroni correction for multiple testing

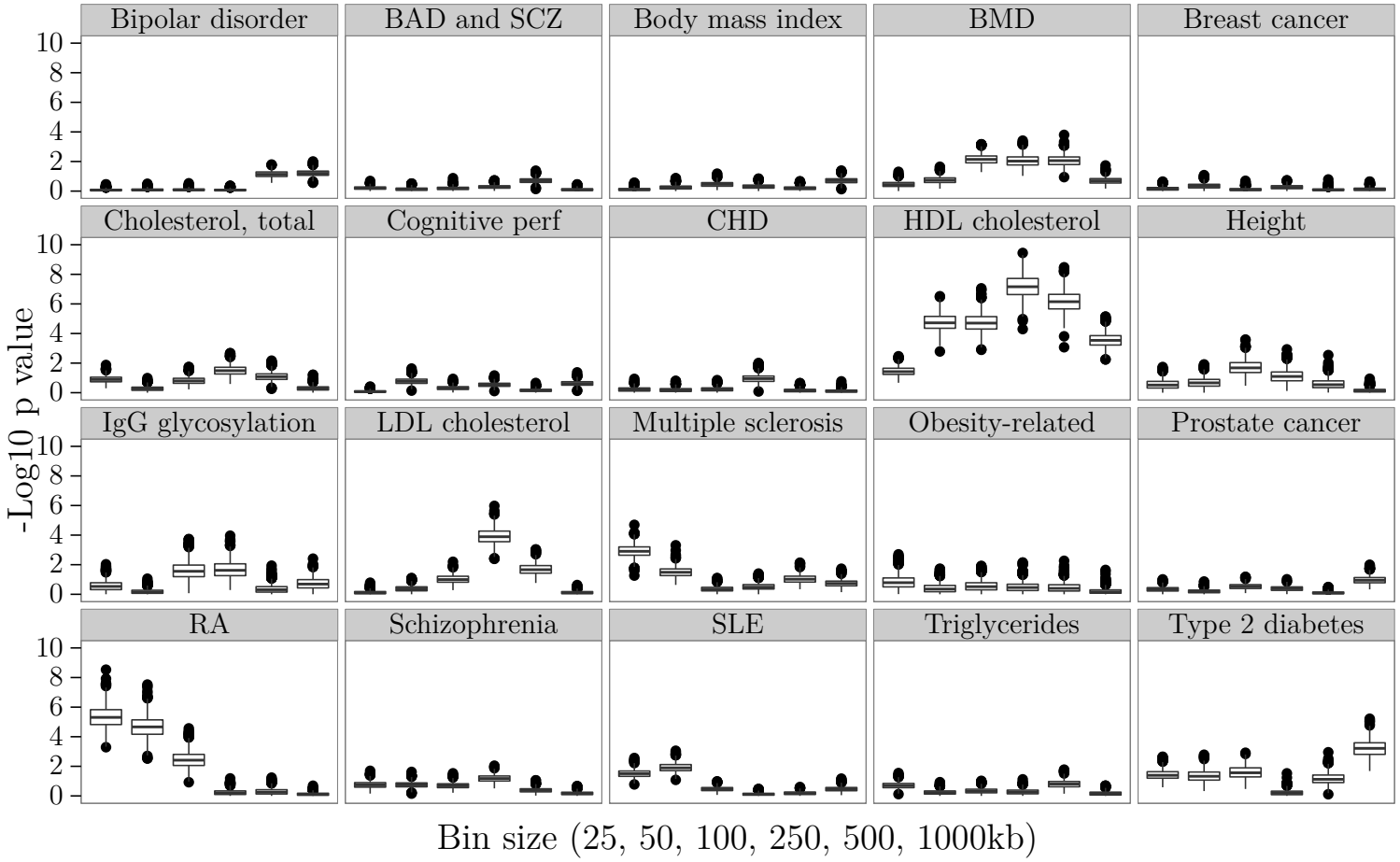


Figure 3.35: Correlation between CD-associated methylation changes and GWAS SNPs for other traits I, see page 106

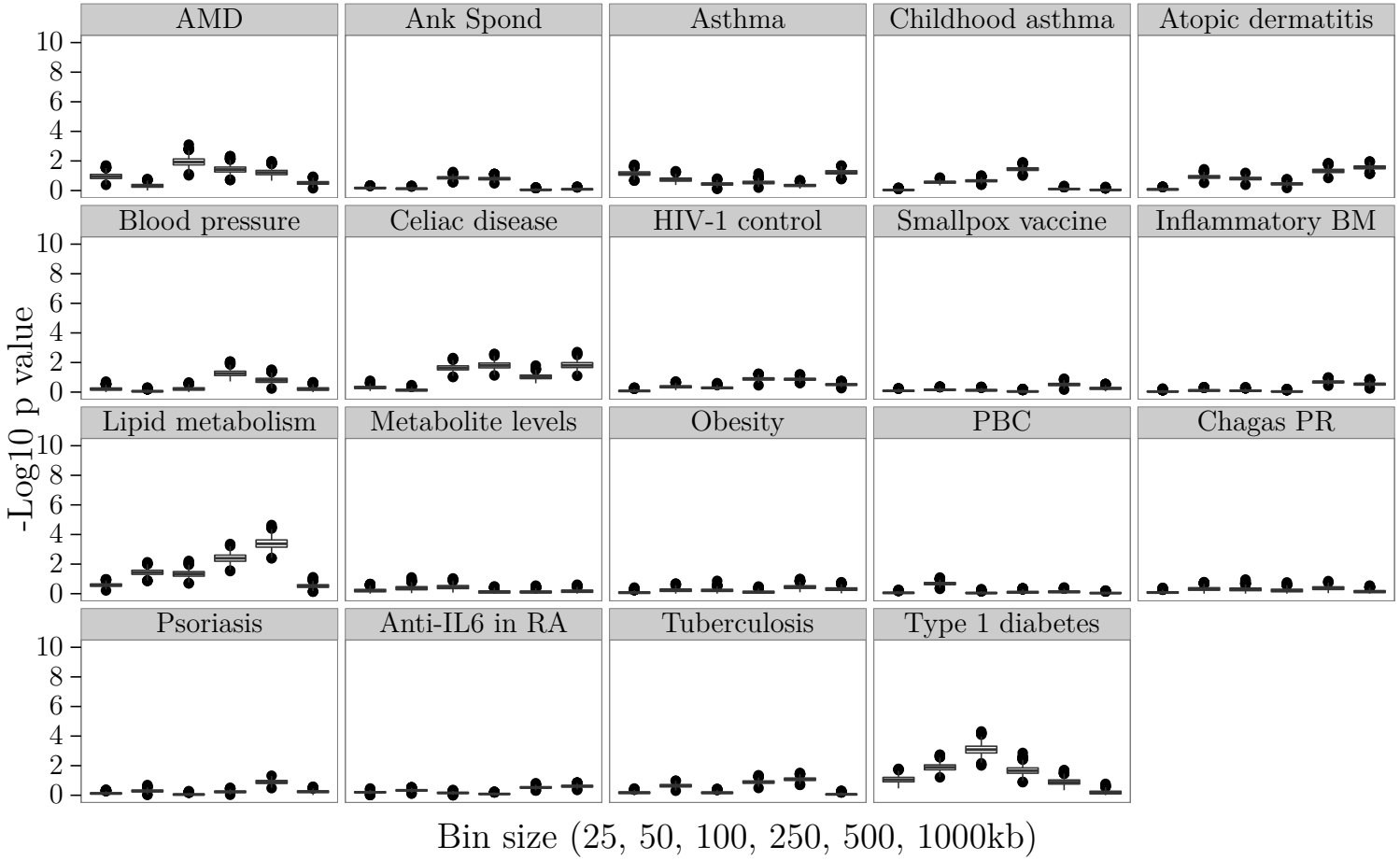


Figure 3.36: Correlation between CD-associated methylation changes and GWAS SNPs for other traits II, see page 106

Table 3.35 and 3.36: Box and whisker plots showing the results for 1000 replications of comparing methylation within range of GWAS SNPs for well studied traits and various immune and inflammatory conditions with randomly selected non-IBD GWAS SNPs. Whiskers extend up to 1.5 × IQR from the 1st and 3rd quartiles, with any remaining results plotted as points.

BAD and SCZ: Bipolar affective disorder and schizophrenia, **BMD**: Bone mineral density, **Cognitive perf**: cognitive performance, **CHD**: Coronary heart disease, **RA**: Rheumatoid arthritis, **SLE**: Systemic lupus erythematosus, **AMD**: Age-related macular degeneration, **Ank Spond**: Ankylosing spondylitis, **Smallpox vaccine**: IL-6 response to smallpox vaccination, **Inflammatory BM**: Inflammatory biomarkers, **PBC**: Primary biliary cirrhosis, **Chagas PR**: PR interval in *Tripanosoma cruzi* seropositivity, **Anti-IL6 in RA**: Response to tocilizumab in rheumatoid arthritis.

3.3.6 GO term analysis

Genes containing more probes are more significantly likely to have a probe with an FDR adjusted p<0.05 (Spearman's ρ=0.71, p=6.8×10^{-7}, figure 3.37, page 107). The same pattern is not seen for the small number of genes containing Bonferroni significant probes (Spearman's ρ=0.53, p=0.12, figure 3.38, page 108). GO term analysis was performed on genes containing one or more FDR significant probes after applying a correction based on the number of probes annotated to that gene as described by Geeleher *et al.* [207].

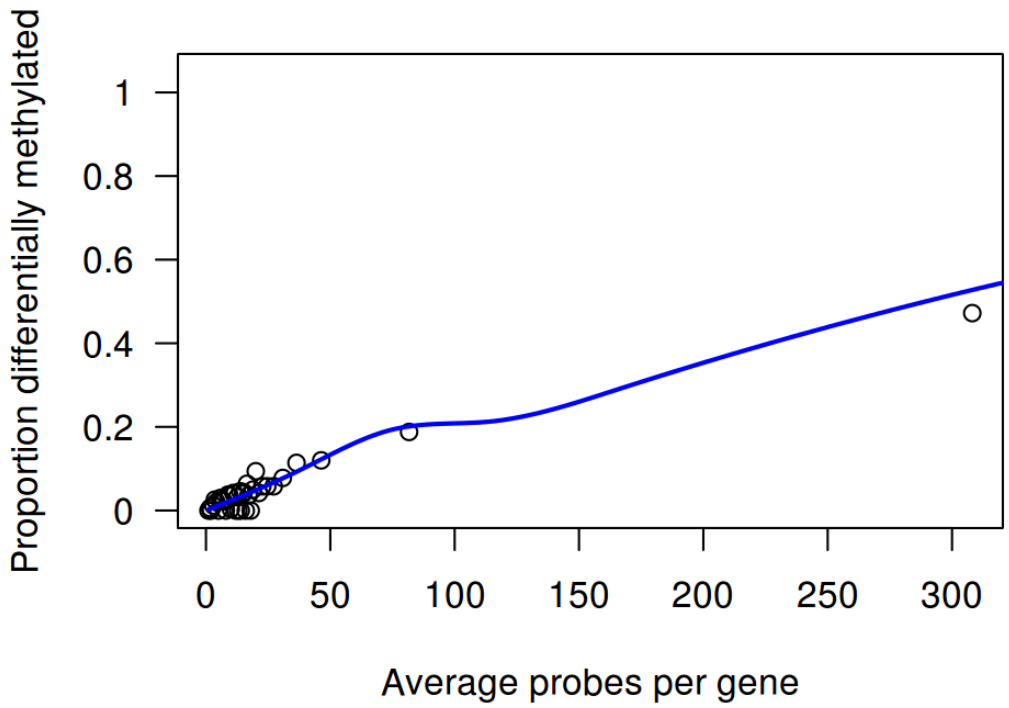


Figure 3.37: Genes assayed by more probes are increasingly likely to demonstrate significant methylation differences at one probe. Spearmans's ρ=0.71, p=6.8×10^{-7}

Table 3.13 shows the most significantly enriched GO terms amongst probes with FDR significant methylation differences and table 3.14 shows the GO terms which were enriched among Bonferroni significant probes. The p values for GO term enrichment are corrected by the Benjamini-Hochberg method for multiple testing.

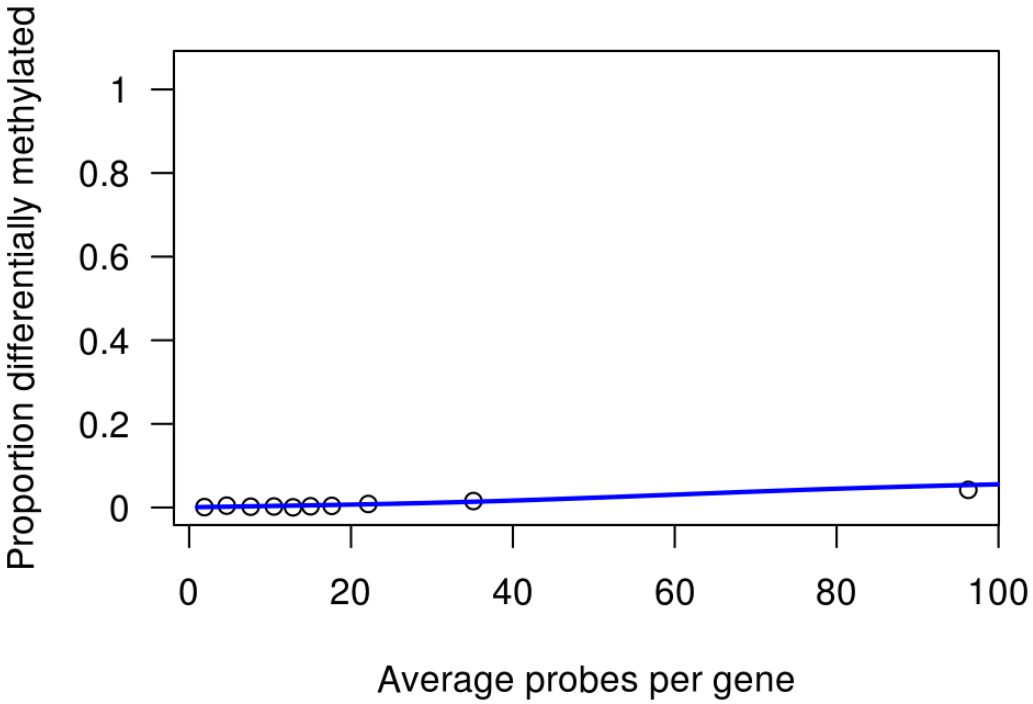


Figure 3.38: No bias to increasing likelihood of Bonferroni significant methylation differences with increasing probe number. Spearman's ρ=0.53, p=0.12

GO Term	P value	Description
GO:0048583	$6.86\text{x}10^{-11}$	regulation of response to stimulus
GO:0002757	$6.86\text{x}10^{-11}$	immune response-activating signal transduction
GO:0048584	$6.86\text{x}10^{-11}$	positive regulation of response to stimulus
GO:0002253	$9.59\text{x}10^{-11}$	activation of immune response
GO:0002764	$3.96\text{x}10^{-10}$	immune response-regulating signaling pathway
GO:0007165	$5.19\text{x}10^{-10}$	signal transduction
GO:0050778	$5.46\text{x}10^{-10}$	positive regulation of immune response
GO:0023052	$5.53\text{x}10^{-10}$	signaling
GO:0044700	$5.53\text{x}10^{-10}$	single organism signaling
GO:0007154	$9.93\text{x}10^{-10}$	cell communication
GO:0050896	$2.04\text{x}10^{-9}$	response to stimulus
GO:0051716	$2.04\text{x}10^{-9}$	cellular response to stimulus
GO:1902531	$2.40\text{x}10^{-9}$	regulation of intracellular signal transduction
GO:0035556	$3.46\text{x}10^{-9}$	intracellular signal transduction
GO:0045089	$1.23\text{x}10^{-8}$	positive regulation of innate immune response
GO:0031349	$1.26\text{x}10^{-8}$	positive regulation of defense response
GO:0009966	$1.34\text{x}10^{-8}$	regulation of signal transduction
GO:0002684	$2.06\text{x}10^{-8}$	positive regulation of immune system process
GO:0002224	$2.06\text{x}10^{-8}$	toll-like receptor signaling pathway
GO:0002218	$2.46\text{x}10^{-8}$	activation of innate immune response

Table 3.13: The most significantly enriched GO terms among probes with FDR significant disease-associated methylation changes after correction for probe number. P values corrected for multiple testing (Benjamini-Hochberg FDR).

GO Term	P value	Description
GO:0002252	$6.18x10^{-6}$	immune effector process
GO:0006659	$7.95x10^{-6}$	phosphatidylserine biosynthetic process
GO:0045414	$1.09x10^{-5}$	regulation of interleukin-8 biosynthetic process
GO:0002460	$1.27x10^{-5}$	adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains
GO:0042228	$1.91x10^{-5}$	interleukin-8 biosynthetic process
GO:0006955	$1.97x10^{-5}$	immune response
GO:2001235	$2.20x10^{-5}$	positive regulation of apoptotic signaling pathway
GO:0002455	$3.06x10^{-5}$	humoral immune response mediated by circulating immunoglobulin
GO:2001238	$3.57x10^{-5}$	positive regulation of extrinsic apoptotic signaling pathway
GO:0016064	$3.83x10^{-5}$	immunoglobulin mediated immune response
GO:0002250	$3.94x10^{-5}$	adaptive immune response
GO:0019724	$4.12x10^{-5}$	B cell mediated immunity
GO:0050778	$4.22x10^{-5}$	positive regulation of immune response
GO:0002376	$6.31x10^{-5}$	immune system process
GO:0002718	$1.80x10^{-4}$	regulation of cytokine production involved in immune response
GO:0002682	$1.95x10^{-4}$	regulation of immune system process
GO:0002253	$1.97x10^{-4}$	activation of immune response
GO:0045416	$2.23x10^{-4}$	positive regulation of interleukin-8 biosynthetic process
GO:0002684	$2.23x10^{-4}$	positive regulation of immune system process
GO:0050776	$2.28x10^{-4}$	regulation of immune response

Table 3.14: The most significantly enriched GO terms among probes with Bonferroni significant disease-associated methylation differences after correction for the number of probes per gene. P values corrected for multiple testing (Benjamini-Hochberg FDR).

3.3.7 Enrichment of DNA motifs

DNA motifs with high similarity to human transcription factor binding sites found within 2kb of Illumina 450k probes with Bonferroni significant methylation changes, FDR significant methylation changes and within DMRs are shown in tables 3.24, 3.25, and 3.26 respectively (from page 174). Motifs present in a large proportion of the examined sequences were selected for further investigation. A prevalence threshold of greater than twice the mean prevalence rate of all discovered motifs was used. These are shown in bold in tables 3.16 and 3.17 (pages 112–113), however some of the results with the lowest similarity to transcription factors and lowest prevalence in the input sequences have been omitted to shorten the tables. There was no significant difference in the prevalence of these motifs between Bonferroni significant regions and FDR significant regions by χ^2 test, and within regions containing Bonferroni and FDR significant probes there was no significant correlation between p value and the presence of the motif by Wilcoxon rank-sum test. In table 3.15 the prevalences were also compared to those that would be obtained from a sequence of random DNA bases, and the prevalences observed by repeatedly (n=100) randomly selecting 1000 non significant probes, and constructing non-overlapping regions as described previously.

Motif	Bonferroni[a]	FDR[b]	Random[c]	Control[d]
AVCBTGTTT	24.6*	17.8	7.0	16.3
CCACGBVA	23.1	35.6	28.0	35.8
CCCYTAG	30.8	40.5	25.0	40.7
CTDTGNGGT	30.8	28.1**	9.4	22.8

Table 3.15: The percentage prevalence of each motif within regions containing [a]Bonferroni significant probes, [b]FDR significant probes, [c]random DNA, and [d]regions containing randomly selected non-significant probes. χ^2 test versus control *p=0.038, **p=2.8×10^{-5}

Although they initially appeared to be highly prevalent, CCACGBVA and CCCYTAG are found at a very similar rate in proximity to non-significant methylation probes. AVCBTGTTT is modestly over-represented near Bonferroni significant probes, but the strongest finding is a 23–35% relative enrichment of CTDGNGGT compared to non-significant methylation probes. The p value for enrichment of CTDGNGGT in proximity to Bonferroni significant probes is 0.067, reflecting the small number of regions, though the trend matches that seen in the FDR significant probes.

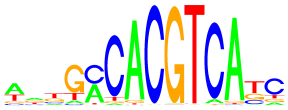
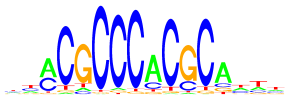
	
CCACGBVA	CREB3L1, CREB3
	
	XBP1
	
	EGR4, EGR3
	
CTDTGNGGT	GLI2
	
reverse complement	RUNX3

Table 3.16: Enriched DNA motifs (left) and matching transcription factor binding sites (right) I

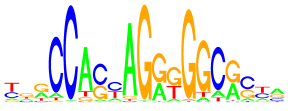
	
CCCYTAG	TFAP2B, TFAP2C, TFAP2A
	
reverse complement	PLAG1
	
	MZF1
	
	CTCF
	
AVCBTGTTT	FOXA1

Table 3.17: Enriched DNA motifs (left) and matching transcription factor binding sites (right) II

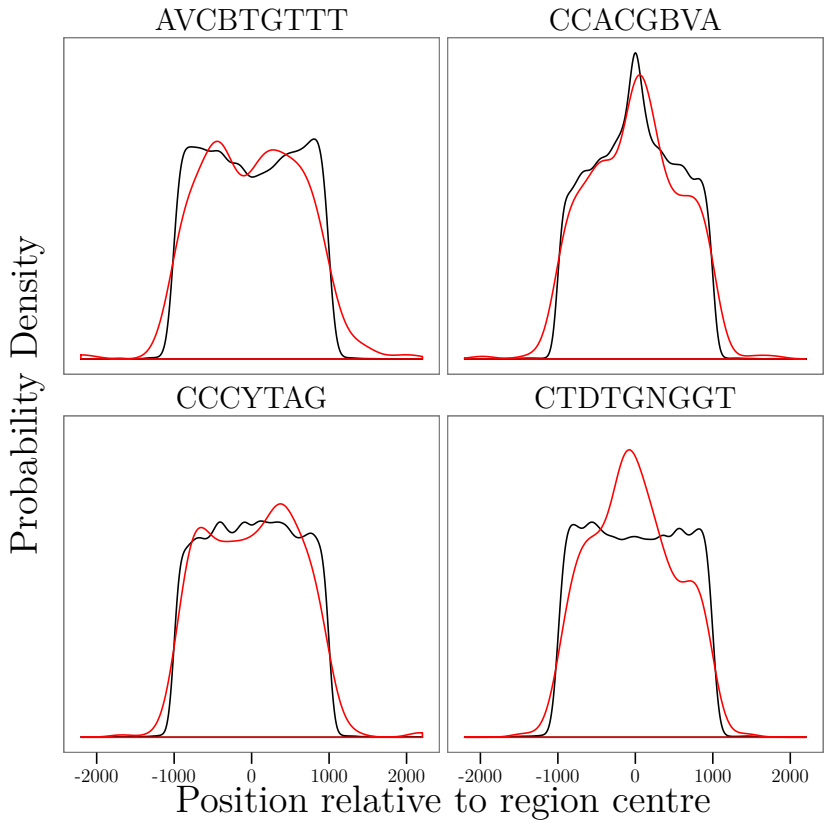


Figure 3.39: Position of proposed motifs relative to methylation probe in 100 random selections of 1000 non-significant probes ∼ and FDR significant probes ∼

In addition to the prevalence relative to that seen in non-significant probes, the positions of the proposed motifs relative to each probe was examined and compared to the position seen in the 100 selections of 1000 random non-significant probes. `AVCBTGTTT` and `CCYTAG` were uniformly distributed across the regions, whereas `CCACGBVA` and `CTDTGNGGT` were more likely to be found close to the probe. In many cases the target of the probe was the CpG contained within the `CCACGBVA` motif, or the motif was found within the probe sequence (table 3.23, page 173). However, this same pattern was observed in non-significant probes. Only `CTDTGNGGT` demonstrated a different distribution pattern near significant methylation differences compared to non-significant probes. The context of the regions containing Bonferroni significant probes and the `CTDGNGGT` motif are shown in table 3.18

Gene	Chr	Sequence	Position	Context
VMP1	17	CTTTGGGGT	+ 694–702	ℝ§ intron
RPS6KA2	6	CTTTGGGGT	- 64–72	ℝ§ intron
	3	CTTTGGGGT	+ 108–116	ℝ no gene
	15	CTGTGAGGT	- 398–406	ℝ no gene
MYO1E	15	CTTTGGGGT	- 642–650	ℝ intron
	12	CTTTGGGGT	- 94–102	ℝ no gene
	2	CTTTGAGGT	+ 350–358	ℝ no gene
CDC42BPB	14	CTGTGAGGT	+ 886–894	ℝ intron
RUNX3	1	CTTTGAGGT	- 796–804	ℝ intron
MPRIP	17	CTGTGGGGT	- 671–679	ℝ§ intron
SLC10A6	4	CTGTGGGGT	- 646–654	intron
NMRAL1	16	CTGTGGTGT	+ 277–285	ℝ exon
SYNJ2	6	CTGTGGGGT	- 298–306	ℝ§ intron
PTDSS2	11	CTGTGGGGT	- 551–559	ℝ intron
FKBP5	6	CTATGAGGT	+ 337–345	ℝ intron
TRPS1	8	CTATGGGGT	+ 520–528	ℝ intron
	15	CTTTGAGGT	- 440–448	ℝ no gene
	15	CTGTGAGGT	- 857–865	ℝ no gene
	20	CTGTGAGGT	+ 949–957	§ 5′/3′/intron
TBPL1	6	CTGTGTGGT	- 101–109	ℝ intron

Table 3.18: Context of the CTDGNGGT motif near Bonferroni significant probes. Position in bp relative to nominal probe target. ℝ: Regulatory region (DNA I hypersensitivity, H3K27ac, and transcription factor binding from ENCODE), § possible alternative promoter region.

3.3.8 Replication by pyrosequencing

Pyrosequencing of BISCUIT DNA samples

Figure 3.40 shows pyrosequencing analysis of the most significant CpG in a separate bisulfite conversion of the DNAs from the paediatric discovery cohort. This reproduced the results with an r^2 of 0.95. Further CpGs were not replicated in the paediatric DNA because the Illumina 450k platform has been extensively validated and the paediatric DNA was extremely limited.

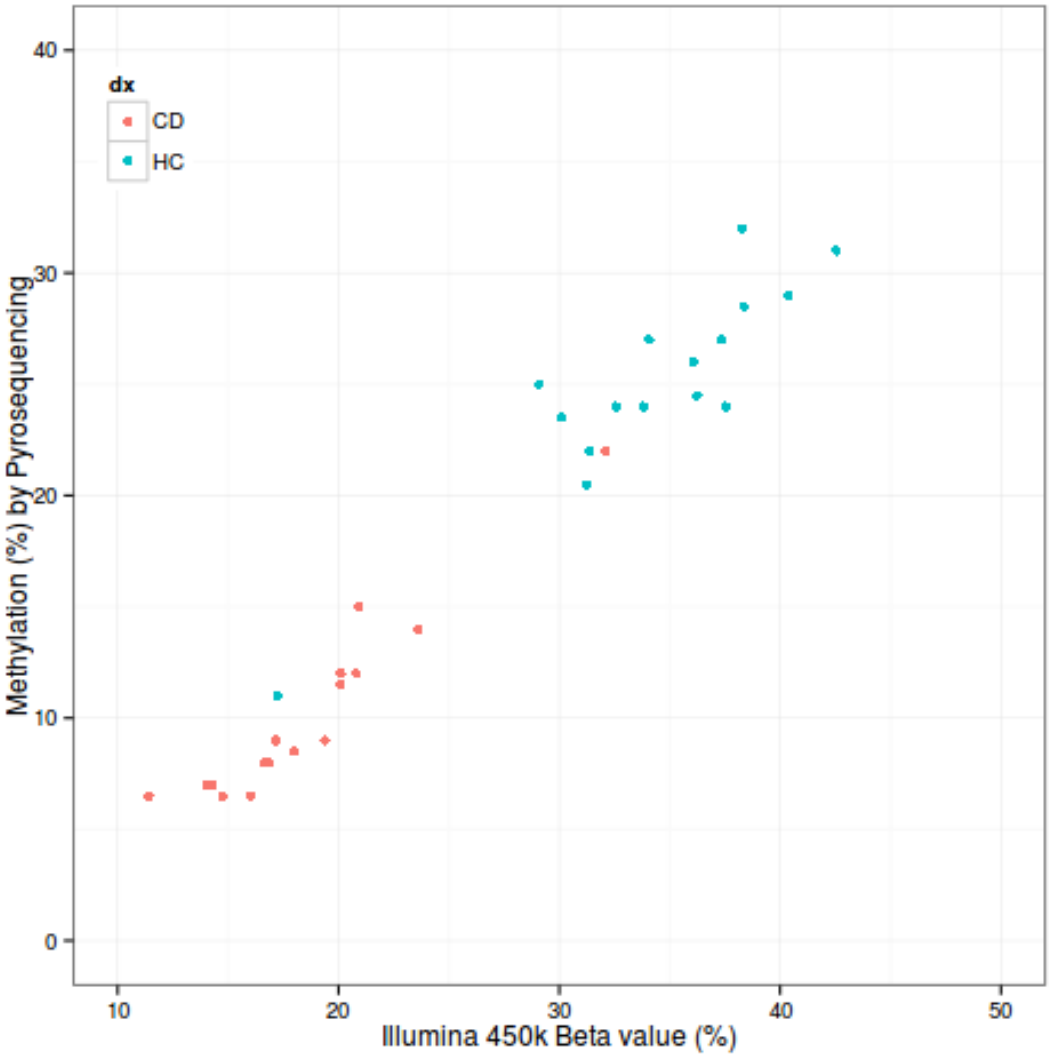


Figure 3.40: Illumina 450k beta values for cg12054453, and pyrosequencing methylation values for the same CpG (r^2 0.95)

Adult replication by pyrosequencing

Probes were selected for replication by pyrosequencing after the initial analysis of the paediatric Illumina 450k data. This primary analysis used the ratio of neutrophils to lymphocytes in the samples which had full blood counts taken the same day as the DNA sample as a covariate, rather than the methylation-derived full blood counts.

Probes were chosen based on their significance in this initial analysis, links with relevant biological pathways or known significance to Crohn's disease, and ease of designing pyrosequencing assays which measure the same CpG as the Illumina 450k probe. This resulted in pyrosequencing assays for *VMP1*, *RPS6KA2*, *SOCS3*, *ARHGEF3*, *SBNO2*, *CFI*, *CDC42BPB* and *TOLLIP*. Of these probes, only *TOLLIP* lost significance in the final Illumina 450k analysis, where an deconvoluted cell compositions were used as covariates rather than neutrophil to leucocytes ratio.

In a cohort of 20 adults with Crohn's disease and 20 healthy controls, significant differences were found in every tested probe (figure 3.41) except for *TOLLIP* (not shown). Larger scale replications of *VMP1* and *RPS6KA2* were undertaken and are shown in sections 4.3.1 and 5.3.1.

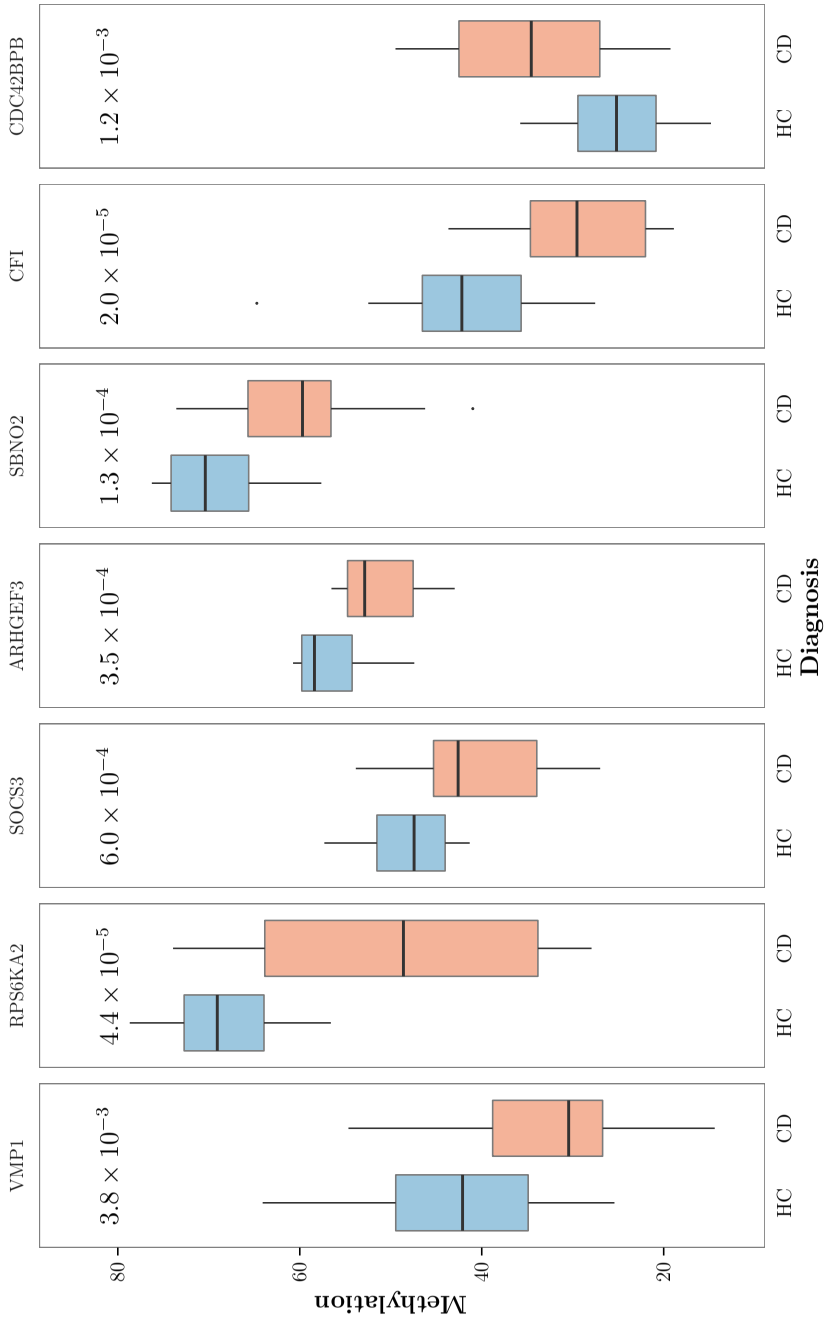


Figure 3.41: Box and whisker plot showing the replication of 7 probes by pyrosequencing in an adult cohort of 20 CD patients and 20 healthy controls. Whiskers extend up to 1.5 × IQR from the 1st and 3rd quartiles, with any remaining results plotted as points.

3.4 Discussion

3.4.1 Implications for future study design

Power

Rakyan *et al.* [509] simulated epigenome-wide association studies in mixed cell populations. Their calculations were based on methylation odds ratio ('methOR'), calculated as shown in figure 3.42 (page 120). The power calculations were based on three scenarios, as shown in figure 3.43 (page 121). In scenarios a and b, control samples exhibited full methylation and case samples (i–iv) showed increasing proportions of unmethylated and partially methylated DNA respectively. In scenario c there is a small hypomethylation and increase in variance. The resulting power to detect the methylation change described by the three scenarios are shown in table 3.19 (page 120). The authors conclude that:

> When methOR is around 1.25, a sample size of 800 cases + 800 controls is adequate to achieve 80% power at a significance level of α=10^{-6} for scenario c but not a or b [...] When methOR is around 1.5, a sample size of 400 + 400 gives 80% power at α=10^{-6} for b and c but not a.
>
> If we regard 1.5 to be a target methOR value, then it would not seem to be cost-effective to pursue an EWAS with fewer than 400 cases and 400 controls, and 800 of each would be preferable to achieve good power [509].

As shown in section 3.3.2, the group sizes necessary to detect results matching those attaining Bonferroni significance in this dataset are much lower. Table 3.20 (pages 122-125) shows the methOR and md of each probe for comparison with the simulated results from Rakyan *et al.* (table 3.19), but with powers for group sizes from 25–100 instead of 100–800. This table also shows the group size needed for 80% power at a significance level of 1.1×10^{-7} (Bonferroni significance).

The methylation changes seen in Bonferroni significant probes most closely match scenario c from Rakyan *et al.* The methylation profiles of the six probes with the smallest and largest group size needed to discover are shown in figures 3.44 and 3.45 respectively (pages 126 and 127). The observed methORs in table 3.20 range from 1.21–1.91 (median 1.38), which match the predicted methORs from table 3.19. The mean differences are however much greater, ranging from 2.74–15.65 (median 7.16).

		(Methylation) Outcome	
		+	-
Exposure (Disease)	CD	meth	1-meth
	HC	meth	1-meth

$$methOR = \frac{meth_{CD} \times (1 - meth_{HC})}{meth_{HC} \times (1 - meth_{CD})}$$

Figure 3.42: The calculation of methylation odds ratio ('methOR') as per Rakyan *et al.* [509]

			Group Size							
			100		200		400		800	
		α	10^{-6}	10^{-8}	10^{-6}	10^{-8}	10^{-6}	10^{-8}	10^{-6}	10^{-8}
	MethOR	md								
	Scenario a									
i.	1.24	3.6	0	0	0	0	1	0	20	3
ii.	1.49	7.2	0	0	4	0	66	21	100	99
iii.	1.78	10.8	2	0	55	11	100	98	100	100
iv.	2.10	14.4	18	1	99	78	100	100	100	100
	Scenario b									
i.	1.24	3.6	0	0	0	0	2	0	33	8
ii.	1.49	7.2	1	0	16	2	85	51	100	100
iii.	1.78	10.8	13	1	84	46	100	100	100	100
iv.	2.10	14.4	60	19	100	97	100	100	100	100
	Scenario c									
i.	1.27	1.25	1	0	7	1	50	19	98	88
ii.	1.54	2.5	37	10	95	78	100	100	100	100
iii.	1.82	3.75	95	77	100	100	100	100	100	100
iv.	2.11	5.0	100	99	100	100	100	100	100	100

Table 3.19: Simulated power for finding methylation differences between groups of 100, 200, 400, or 800 at significance levels of 10^{-6} and 10^{-8}. From Rakyan *et al.* [509].

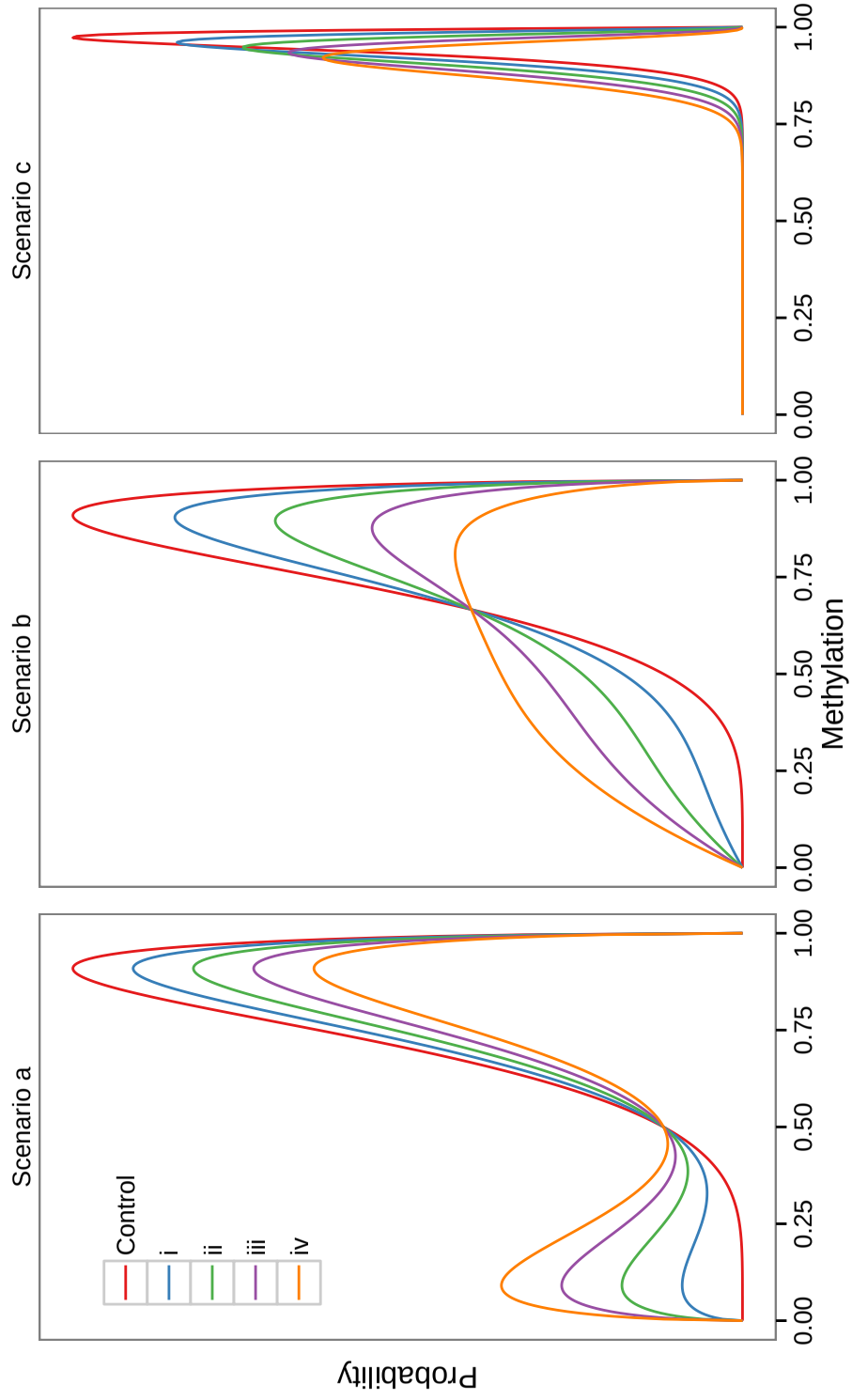


Figure 3.43: Predicted methylation profiles in Rakyan *et al.* [509]

						Group Size:	25		50		75		100	
						α:	10^{-6}	10^{-8}	10^{-6}	10^{-8}	10^{-6}	10^{-8}	10^{-6}	10^{-8}
probe	n	$\bar{\beta}_{cd}$	$\bar{\beta}_{hc}$	MethOR	md		Power							
cg12054453	17	0.23	0.37	1.91	13.39		99.95	98.05	100	100	100	100	100	100
cg17501210	22	0.69	0.79	1.69	9.89		97.57	78.81	100	100	100	100	100	100
cg12992827	21	0.58	0.71	1.77	12.91		98.07	81.37	100	100	100	100	100	100
cg16936953	20	0.39	0.55	1.88	15.65		99.31	89.65	100	100	100	100	100	100
cg04389058	22	0.60	0.68	1.36	7.14		97.05	76.45	100	100	100	100	100	100
cg02560388	24	0.24	0.29	1.33	5.57		92.88	62.64	100	99.99	100	100	100	100
cg00382138	23	0.48	0.58	1.48	9.71		96.33	73.49	100	100	100	100	100	100
cg01671681	27	0.68	0.74	1.32	5.64		86.71	49.50	100	99.92	100	100	100	100
cg19821297	31	0.47	0.58	1.54	10.79		72.58	30.93	99.97	99.14	100	100	100	100
cg27049094	22	0.55	0.46	1.40	8.38		97.21	77.15	100	100	100	100	100	100
cg27310092	23	0.47	0.54	1.29	6.36		95.00	68.82	100	100	100	100	100	100
cg01059398	29	0.44	0.52	1.34	7.36		78.58	37.60	99.99	99.62	100	100	100	100
cg18181703	28	0.42	0.51	1.43	8.91		80.83	40.48	99.99	99.73	100	100	100	100
cg01726890	30	0.81	0.77	1.30	4.26		74.16	32.56	99.98	99.29	100	100	100	100
cg18942579	26	0.50	0.60	1.52	10.29		88.64	53.07	100	99.95	100	100	100	100
cg22768358	26	0.82	0.78	1.34	4.67		88.24	52.30	100	99.94	100	100	100	100
cg02716826	31	0.42	0.47	1.25	5.46		70.51	28.93	99.96	98.90	100	100	100	100
cg20995564	27	0.57	0.69	1.74	12.81		86.87	49.78	100	99.92	100	100	100	100
cg21328643	32	0.82	0.75	1.52	7.16		67.18	25.95	99.93	98.42	100	100	100	100

Table 3.20 Continued overleaf

						Group Size:	25		50		75		100	
						α:	10^{-6}	10^{-8}	10^{-6}	10^{-8}	10^{-6}	10^{-8}	10^{-6}	10^{-8}
probe	n	$\bar{\beta}_{cd}$	$\bar{\beta}_{hc}$	MethOR	md		Power							
cg09349128	27	0.25	0.33	1.50	8.33		84.18	45.33	100	99.85	100	100	100	100
cg18376497	28	0.31	0.40	1.49	9.00		82.73	43.15	100	99.81	100	100	100	100
cg05316065	32	0.56	0.62	1.29	6.18		68.48	27.08	99.94	98.63	100	100	100	100
cg09588020	32	0.46	0.52	1.24	5.41		65.60	24.64	99.91	98.15	100	100	100	100
cg24632582	35	0.50	0.55	1.25	5.63		56.75	18.27	99.71	95.93	100	99.99	100	100
cg08423142	34	0.82	0.85	1.27	3.37		61.44	21.45	99.84	97.27	100	100	100	100
cg19653417	33	0.67	0.74	1.38	6.66		62.44	22.18	99.86	97.50	100	100	100	100
cg07398517	26	0.51	0.58	1.32	6.90		87.52	50.95	100	99.93	100	100	100	100
cg16292768	38	0.52	0.62	1.49	9.81		47.16	12.84	99.17	91.70	100	99.94	100	100
cg18608055	47	0.50	0.62	1.59	11.36		25.49	4.53	93.34	69.62	99.91	98.40	100	99.98
cg17411016	32	0.43	0.50	1.32	6.83		68.39	27.00	99.94	98.61	100	100	100	100
cg27209729	35	0.61	0.68	1.32	6.29		56.18	17.91	99.69	95.74	100	99.99	100	100
cg20782117	32	0.48	0.57	1.43	8.90		65.97	24.94	99.91	98.22	100	100	100	100
cg27361520	27	0.88	0.86	1.28	2.74		86.24	48.68	100	99.91	100	100	100	100
cg24531955	30	0.30	0.36	1.30	5.84		73.26	31.62	99.97	99.21	100	100	100	100
cg24310395	43	0.73	0.65	1.45	7.94		33.29	6.99	96.76	80.24	99.98	99.49	100	100
cg24231380	35	0.83	0.74	1.71	8.86		56.32	18.00	99.70	95.79	100	99.99	100	100
cg00506299	32	0.64	0.73	1.52	9.01		67.90	26.57	99.93	98.54	100	100	100	100
cg01409343	37	0.62	0.70	1.42	7.86		50.94	14.82	99.45	93.64	100	99.97	100	100

Table 3.20 Continued overleaf

						Group Size:	25		50		75		100	
						α:	10^{-6}	10^{-8}	10^{-6}	10^{-8}	10^{-6}	10^{-8}	10^{-6}	10^{-8}
probe	n	$\bar{\beta}_{cd}$	$\bar{\beta}_{hc}$	MethOR	md		Power							
cg12170787	51	0.51	0.59	1.42	8.70		20.50	3.23	89.45	60.62	99.74	96.69	100	99.91
cg11821200	39	0.61	0.56	1.22	4.76		44.96	11.77	98.96	90.37	100	99.92	100	100
cg04987734	28	0.49	0.41	1.38	7.96		82.03	42.14	100	99.78	100	100	100	100
cg27058497	38	0.60	0.50	1.50	9.96		46.73	12.63	99.13	91.45	100	99.94	100	100
cg13134297	39	0.31	0.35	1.21	4.28		44.41	11.51	98.90	90.01	100	99.91	100	100
cg05673882	52	0.39	0.44	1.24	5.28		18.48	2.76	87.28	56.40	99.6	95.54	100	99.85
cg02734358	35	0.40	0.46	1.28	6.05		57.20	18.56	99.73	96.08	100	99.99	100	100
cg10472711	35	0.59	0.67	1.41	8.02		55.42	17.43	99.66	95.47	100	99.99	100	100
cg23842572	32	0.76	0.68	1.48	7.94		69.07	27.60	99.94	98.71	100	100	100	100
cg06653632	30	0.85	0.80	1.40	4.82		74.62	33.04	99.98	99.33	100	100	100	100
cg18860310	35	0.63	0.73	1.59	10.08		54.85	17.08	99.64	95.27	100	99.98	100	100
cg27023597	50	0.63	0.68	1.27	5.47		21.96	3.59	90.78	63.46	99.81	97.32	100	99.94
cg14880894	42	0.80	0.84	1.25	3.27		36.29	8.09	97.56	83.41	99.99	99.68	100	100
cg01850135	36	0.82	0.74	1.56	7.57		53.40	16.22	99.58	94.71	100	99.98	100	100
cg03940776	34	0.40	0.47	1.32	6.74		60.98	21.12	99.83	97.16	100	100	100	100
cg17618872	36	0.79	0.74	1.35	5.43		51.87	15.34	99.50	94.06	100	99.97	100	100
cg08508455	38	0.63	0.68	1.27	5.42		47.82	13.17	99.23	92.06	100	99.95	100	100
cg07839457	48	0.50	0.60	1.48	9.74		24.37	4.22	92.62	67.77	99.88	98.12	100	99.97
cg09137630	34	0.31	0.36	1.23	4.62		61.53	21.52	99.84	97.29	100	100	100	100

Table 3.20 Continued overleaf

						Group Size:	25		50		75		100	
						α:	10^{-6}	10^{-8}	10^{-6}	10^{-8}	10^{-6}	10^{-8}	10^{-6}	10^{-8}
probe	n	$\bar{\beta}_{cd}$	$\bar{\beta}_{hc}$	MethOR	md		Power							
cg22284398	40	0.24	0.30	1.37	6.15		40.45	9.75	98.37	87.12	100	99.83	100	100
cg03546163	37	0.55	0.63	1.40	8.20		49.16	13.87	99.33	92.78	100	99.96	100	100
cg20477259	36	0.74	0.66	1.42	7.28		52.04	15.43	99.51	94.14	100	99.97	100	100
cg27039118	35	0.64	0.70	1.36	6.83		55.73	17.62	99.68	95.58	100	99.99	100	100
cg26581729	49	0.51	0.55	1.21	4.77		23.03	3.86	91.65	65.44	99.85	97.71	100	99.96
cg04345034	35	0.27	0.34	1.34	6.25		54.90	17.11	99.64	95.28	100	99.98	100	100
cg18157896	39	0.40	0.47	1.37	7.75		45.17	11.87	98.98	90.50	100	99.92	100	100
cg08692676	32	0.53	0.61	1.39	8.02		68.21	26.84	99.94	98.58	100	100	100	100

Table 3.20: MethORs and power to detect Bonferroni significant probes. n is the group size needed for 80% power at α=1.1×10^{-7}, $\bar{\beta}_{cd}$ and $\bar{\beta}_{hc}$ are the mean methylation values for Crohn's and control groups; MethOR, md, and power as per table 3.19

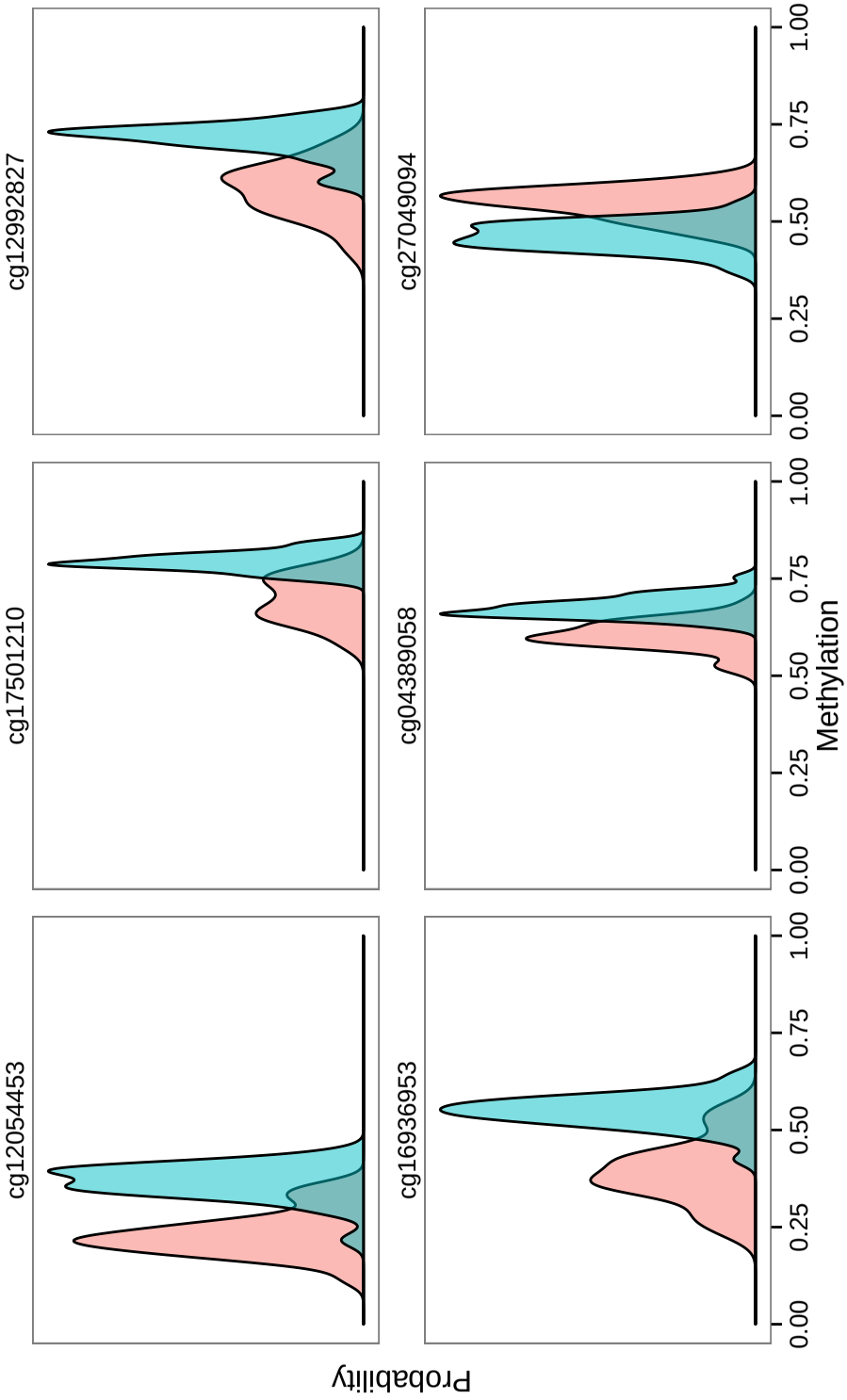


Figure 3.44: Methylation profiles of the six probes with the greatest power as shown in table 3.20

Figure 3.45: Methylation profiles of the six probes with the least power as shown in table 3.20

Visual consideration of the observed patterns of methylation change in Bonferroni significant probes reveals two broad patterns when agnostic to groups and direction, with a small minority showing an intermediate pattern. Examples of probes demonstrating each of these patterns are shown in figure 3.46 (page 130); the changes in mean methylation and standard deviation have been normalized against a standard representative control pattern for ease of comparison.

Pattern	Description
Slip	One end of range unchanged, other end extends Large increase in variance
Slide	Both ends of range shifted small or no change in variance
Intermediate	Both ends of range shifted Medium increase in variance

Table 3.21: Patterns of methylation change

Pattern	DMR_{Top}	DMR_{All}	Bonf	FDR
Slip	52.6	59.5	60.0	49.7
Slide	36.8	35.4	35.4	49.1
Indeterminate	10.5	5.1	4.6	1.1
Total probes	19	79	65	1319

Table 3.22: Percentage of probes matching each pattern of disease associated methylation change for the most significant probe within each DMR (DMR_{Top}), all DMR probes (DMR_{All}), Bonferroni significant probes (Bonf), and FDR significant probes (FDR)

The differences in proportional prevalence of the two main patterns between groups were not statistically significant, however probes exhibiting the 'slip' pattern were slightly more significant ($p=8.02\times10^{-6}$, Spearman's rank correlation $\rho=0.13$), and required slightly smaller group sizes for 80% power at estimated FDR significance levels ($p=9.52\times10^{-5}$, $\rho=0.11$). There was only a weak correlation between methylation pattern and absolute β difference ($p=0.039$, $\rho=0.059$), but the methylation pattern was strongly correlated with β difference including direction ($p=8.03\times10^{-18}$, $\rho=0.24$) and hypomethylation in general (χ^2 $p=1.33\times10^{-16}$, OR:1.56 for hypomethylation with the 'slide' pattern).

The 'slip' pattern is symmetrical — at some probes there is an increase in the variance seen in CD, in others the variance is higher in controls. In Bonferroni significant probes 69.2% of probes exhibiting this pattern have a higher variance in CD, in FDR significant probes the proportion is significantly less (33.7%, χ^2 $p=1.52\times10^{-5}$).

In probes where the mean methylation for all samples was <50%, 88.2% of probes showed hypomethylation in CD, whereas in probes where the mean methylation for all samples was >50%, 67.0% of probes showed hypermethylation in CD (χ^2 p=1.67×10^{-78}). This finding of methylation moving away from 50% in CD compared to controls was seen in both methylation patterns similarly (p=0.065).

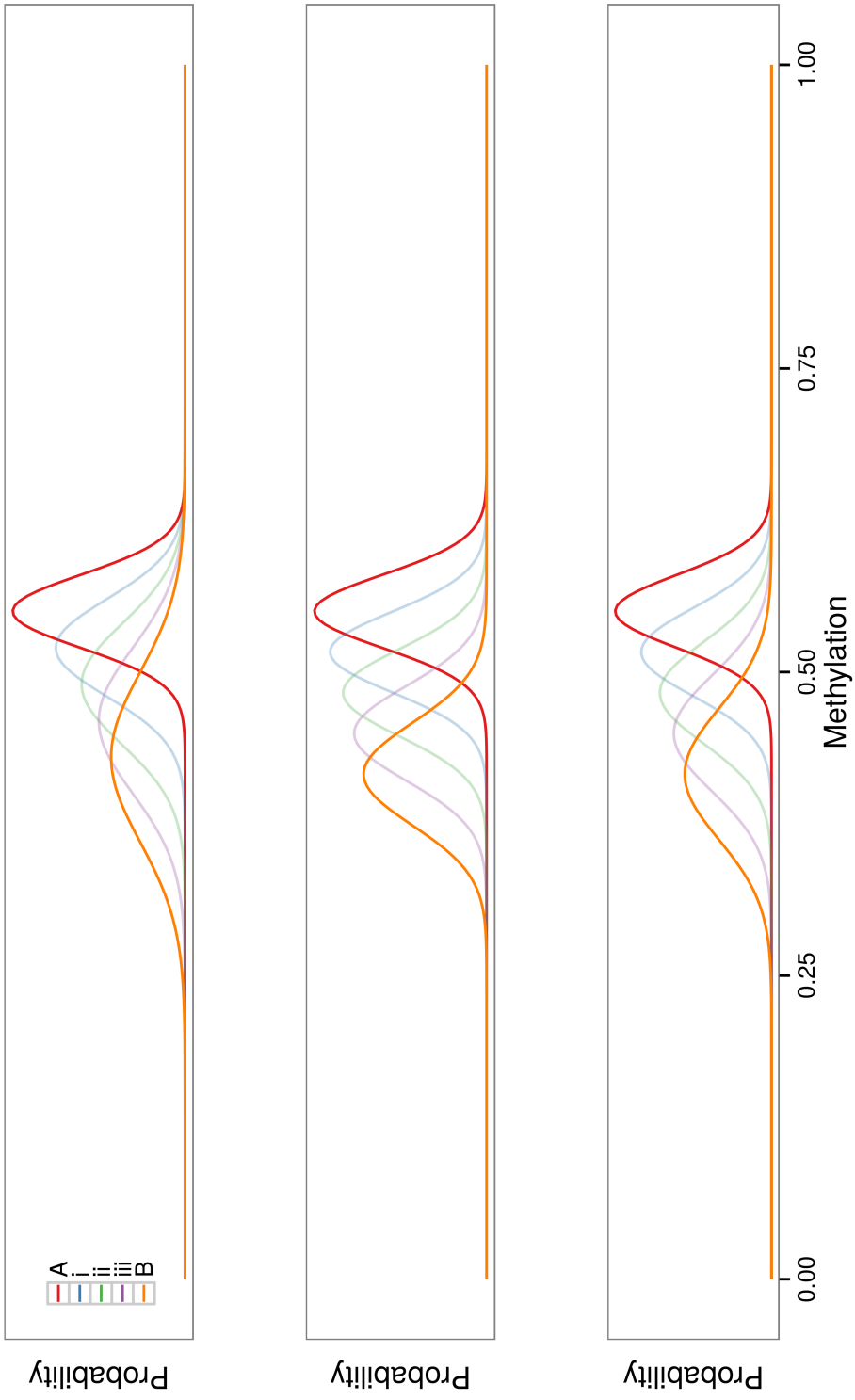


Figure 3.46: Representative examples of the observed patterns of methylation change (orange) normalised against control samples for the top pattern (red), with 3 equidistant weaker results. Top: 'slip' with a large increase in variance, and only a small change at one end of the range. Middle: 'slide' with a large shift in values, but a non-significant change in the range. Bottom: an example of an intermediate result.

Correction for multiple testing

One shortfall of the published EWAS studies in IBD is the inconsistency in the chosen significance thresholds (see table 2.4, page 33) compared with GWAS, where a threshold of 5×10^{-8} is generally accepted. IBD EWAS studies have used methods including the Bonferroni, Holm, and Šidák corrections, the Benjamini-Hochberg false discovery rate (α=0.01, 0.05, or 0.2), and uncorrected p values (<0.05 or 10^{-4}). This, and other varying factors in analysis and design such as cell type correction and the reporting of CpGs or DMRs makes it difficult to see the literature as a cohesive whole, and assess the consistency of results.

The analysis reported in this chapter uses two significance thresholds, The Bonferroni correction and the Benjamini-Hochberg false discovery rate [52]. The justification for this decision and an explanation of the methods used in the IBD EWAS literature follows.

Familywise error rate The Bonferroni method corrects for the familywise error rate (FWER) α, this is the probability of rejecting at least one true null hypothesis. For a set of m p values and associated null hypotheses

$$P_1, P_2, \ldots, P_m$$
$$H_1, H_2, \ldots, H_m$$

Each null hypothesis H_i is rejected if

$$P_i \leq \frac{\alpha}{m}$$

The Šidák (or Dunn-Šidák) correction is similar to the Bonferroni correction [570]. If the probability of a type I error in an individual test is α_t, then the familywise error is α_f

$$\alpha_f = 1 - (1 - \alpha_t)^m$$
$$\therefore \alpha_t = 1 - (1 - \alpha_f)^{\frac{1}{m}}$$

Each null hypothesis H_i is rejected if

$$P_i \leq 1 - (1 - \alpha_f)^{\frac{1}{m}}$$

$$1 - (1 - \alpha_f)^{\frac{1}{m}} \geq \frac{\alpha}{m}$$

and therefore the Šidák correction is less conservative than the Bonferroni correction, however the difference is minimal especially as m becomes large, and would not change the significance of any result in the combined paediatric Illumina 450k dataset.

The Holm method (also called Bonferroni-Holm or Holm-Bonferroni) takes an ordered list of m p values and associated null hypotheses.

$$P_1 \leq P_2 \leq \cdots \leq P_m$$
$$H_1, H_2, \ldots, H_m$$

Then let k be the smallest value of i for which

$$P_i \geq \frac{\alpha}{m + 1 - i}$$

Null hypotheses $H_1 \ldots H_{k-1}$ are rejected and $H_k \ldots H_m$ are accepted. Again, although this is less prone to type II error than the Bonferroni correction, no additional probes would be considered significant in the combined paediatric dataset.

False discovery rate Correction for the false discovery rate is a newer technique than FWER correction. An acceptable false discovery rate is chosen by the experimenter, usually as in this case q=0.05. FDR significance is then calculated in a stepwise manner on the sorted list of p values for an associated list of m null hypotheses.

$$P_1 \leq P_2 \leq \cdots \leq P_m$$
$$H_1, H_2, \ldots, H_m$$

Then let k be the largest i for which

$$P_i \leq \frac{i}{m}q$$

and reject this and all preceding equivalent null hypotheses

$$H_1, H_2, \ldots H_k$$

Thus for the combined paediatric Illumina 450k data, where the total number of probes (m) after removal of those containing SNPs or of poor quality is 449211.

$$\begin{aligned} P_{1319} &\leq \frac{1319}{449211} \times 0.05 \\ 0.0001462 &\leq 0.0001468 \\ P_{1320} &\nleq \frac{1320}{449211} \times 0.05 \\ 0.00014695 &\nleq 0.00014692 \\ \therefore k &= 1319 \end{aligned}$$

This is shown graphically in figure 3.47. By accepting up to 5% false positive results, the number of significant results increases from 65 by any of the FWER methods detailed above to 1319. Benjamini and Hochberg point out in their original paper

> Often the control of the FWER is not quite needed. The control of the FWER is important when a conclusion from the various individual inferences is likely to be erroneous when at least one of them is [...] The overall conclusion that the treatment is superior need not be erroneous even if some of the null hypotheses are falsely rejected [52]

The permutation test described in section 3.3.3 tested the FWER. In 1000 simulated experiments using random permutations of diagnosis with methylation results there were no p values $<2.0\times10^{-7}$ in 95% of the simulated experiments. The FWER correction methods do not account for the correlation observed between neighbouring CpGs. The DMR analysis (section 3.3.4) only identified 19 DMRs from the whole genome, but this understates the correlation between

neighbouring probes. DMRs had to contain at least 3 consecutive probes within a short distance[2] with FDR significant methylation changes between CD and control sharing the same direction of methylation change. This excludes regions where the probe density is too low to be classified as a DMR, where the methylation direction varies, or where there is no disease-associated methylation change, yet the probes are still correlated with each other.

However, there are benefits to using an overly conservative correction as it may only be possible to follow-up a limited number of 'hits'. The approach taken with this work was to correct for the familywise error rate, to produce a smaller list of highly significant results suitable for detailed analysis or functional work, but also a larger list with an acceptable small proportion of expected false discoveries, to allow analysis of wider patterns such as GO term enrichment, colocalisation with genetic risk loci, and defining DMRs. The Holm correction is undoubtedly superior to the Bonferroni correction; however, there were no differences in probe significance between these methods in this dataset and so the p values are presented as Bonferroni corrected because the technique was felt to be more widely known and understood.

Probes per gene As discussed above, the familywise error rate of 0.05 corresponds to a 5% probability of a type I error in the Bonferroni significant probes, and a false discovery rate of 0.05 corresponds to a probability of type I error in 5% of the FDR significant probes. However, much analysis takes place at the level of genes, rather than probes. GO term enrichment and methylation-GWAS correlation is biased by the varying numbers of probes per gene, and an increasing probability of finding a probe with a significant disease-associated methylation difference after FDR correction (but not Bonferroni correction) in genes with more probes (as shown in figure 3.37, page 107).

Although this is recognized [207], and efforts were made to control for this bias, pathway analysis of methylation data remains problematic. Unlike examining expression data for pathway enrichment, methylation changes may not relate to expression, or could act at a distance, and it can not be assumed that methylation would be correlated to the regulation of genes at every level of a pathway.

In the GO term analysis a probability weighting function was used to correct for the expected probability of each gene containing an FDR significant probe, based on the number of probes that gene contained, and in the GWAS colocalisa-

[2]Adjusted for CpG density in various genomic regions

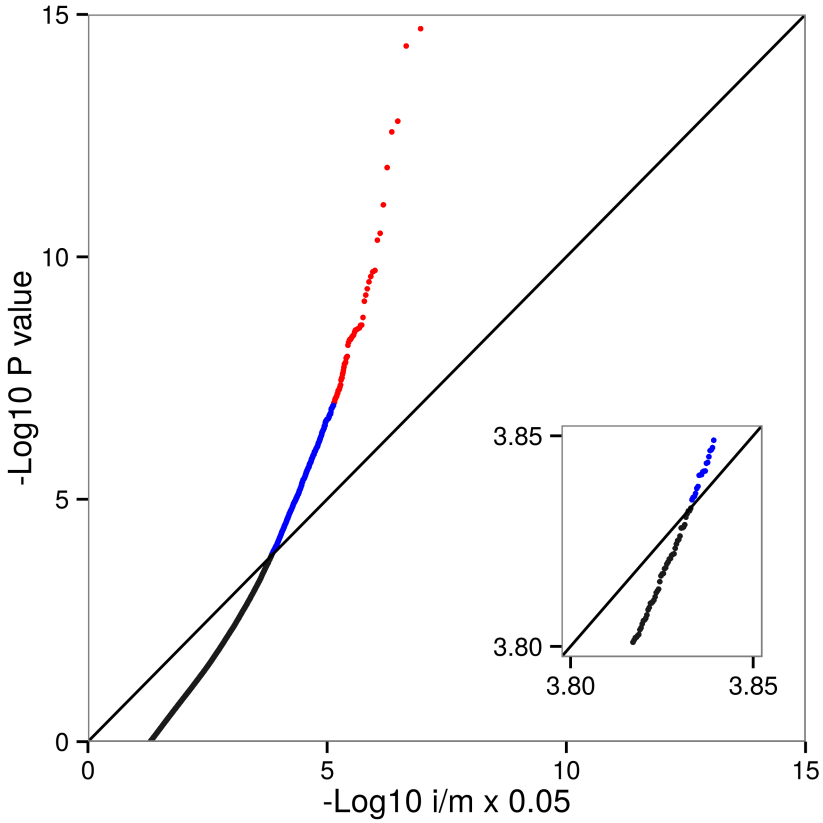


Figure 3.47: -Log_{10} P_i from the combined paediatric Illumina 450k dataset plotted against $^{\text{i}}/_{\text{m}} \times q$, the line is where $P_i = {}^{\text{i}}/_{\text{m}} \times \text{q}$ and therefore all results before the line is crossed are significant. Blue points are FDR significant probes, red points are Bonferroni significant.

tion analysis probe density over genomic regions rather than genes was corrected for. This is more rigorous than our group's previous analytical technique [458]. The number of methylation measurements made by arrays has increased from approximately 27,000 to over 480,000, and the Illumina 450k will be discontinued at the end of 2015 to be replaced by the Illumina Infinium MethylationEPIC Kit which measures over 850,000 CpGs. Concurrently costs of bisulfite NGS are falling. This will make corrections for multiple testing and for variations in probe density increasingly important in epigenome-wide association studies.

3.4.2 β values & M values

The analysis of DNA methylation using the Illumina Infinium range of technologies involves several idiosyncrasies, which must be appreciated in order to produce a dependable analysis, and also in order to understand the limitations of the analysis and the assumptions upon which it is based. The analysis presented in this chapter was performed on β values, whereas many similar analyses are performed using M values. This section will explain the two, demonstrate a minimal impact on results between them, and discuss potential alternatives.

Linear models using β values, and transformed β values

Methylation is reported by the Illumina 450k platform as a β value, which is conceptually equivalent to the proportion of methylated cytosine at a particular base. Where M and U are the methylated and unmethylated signal intensities, and α is a constant (100 for the 450k platform) preventing erroneous values when M and U are small:

$$\beta \ value = \frac{M}{M + U + \alpha}$$

Asymmetry is frequently seen in measures of proportions [184], with β values demonstrating a skewed bimodal distribution (as shown in figure 3.4, page 64). There is frequently a degree of heteroscedasticity — with variance tending to be smaller near extremes of methylation compared to central values, undermining the assumption of homoscedasticity in statistical procedures such as Gaussian regression [690]. Finally, β values are bounded between 0 and 1, which produces non-linear relationships between explanatory variables and β, as a linear relationship would not be bounded between 0 and 1 [675].

The use of logit-transformed β values ("M values") was proposed to deal with the shortcomings in β values [157]. Using a default value of $\alpha = 1$;

$$M \ value = \log_2\left(\frac{M + \alpha}{U + \alpha}\right) \approx \log_2\left(\frac{\beta}{1 - \beta}\right)$$

Du *et al.* [157] did show that M values improved the heteroscedasticity and distribution of β values, however they only considered a single dataset and feature selection was based on fold change rather than test statistics or p value. Trans-

formation to M values only partially deals with heteroscedasticity, and to an even lesser extent, the asymmetrical distribution of methylation data [539]. Zhuang *et al.* [738] performed a more in-depth study of 7 datasets and showed comparable false discovery rates when analysing β values and M values, and only found the use of M values to be advantageous when the number of samples of each phenotype was very small (2–3). The authors also show that supervised PCA with β values outperform M values as "it appears that the correlation structure between biologically relevant CpGs is compromised leading to worse modelling of the biological variation" [738].

Analysis of the combined paediatric dataset (figure 3.48, page 139), as well as the IBD Character dataset[3] (figure 3.49, page 140) show strong correlations between the results obtained in either analysis (ρ=0.92–0.99), with differences in whether a probe is found to be significant limited to a small proportion of probes close to the limit of significance.

M values lack the intuitive meaning of β values, as they can take any value $-\infty$–∞, and the relationship between covariates and methylation can only be considered on the transformed scale [184]. Additionally heteroscedasticity in the data may be biologically relevant [539, 738], and disease-associated changes in methylation variance as well as changes in mean methylation are notable in the results presented in this chapter. Using alternative analytical techniques which do not depend on symmetry and uniform variance, rather than attempting to correct heteroscedasticity may be a superior approach [184].

The arcsine square root transformation is often used in ecological studies to address heteroscedasticity, and has seen some use with methylation results [64, 483], however results remain bounded though between 0 and $\pi/2$ [675], as and with M values it may be a better option to use alternative regression methods, rather than using different transformations to reduce the complexity of the data.

$$A\ value = \sin^{-1}\left(\sqrt{\beta\ value}\right)$$

[3] see appendix A

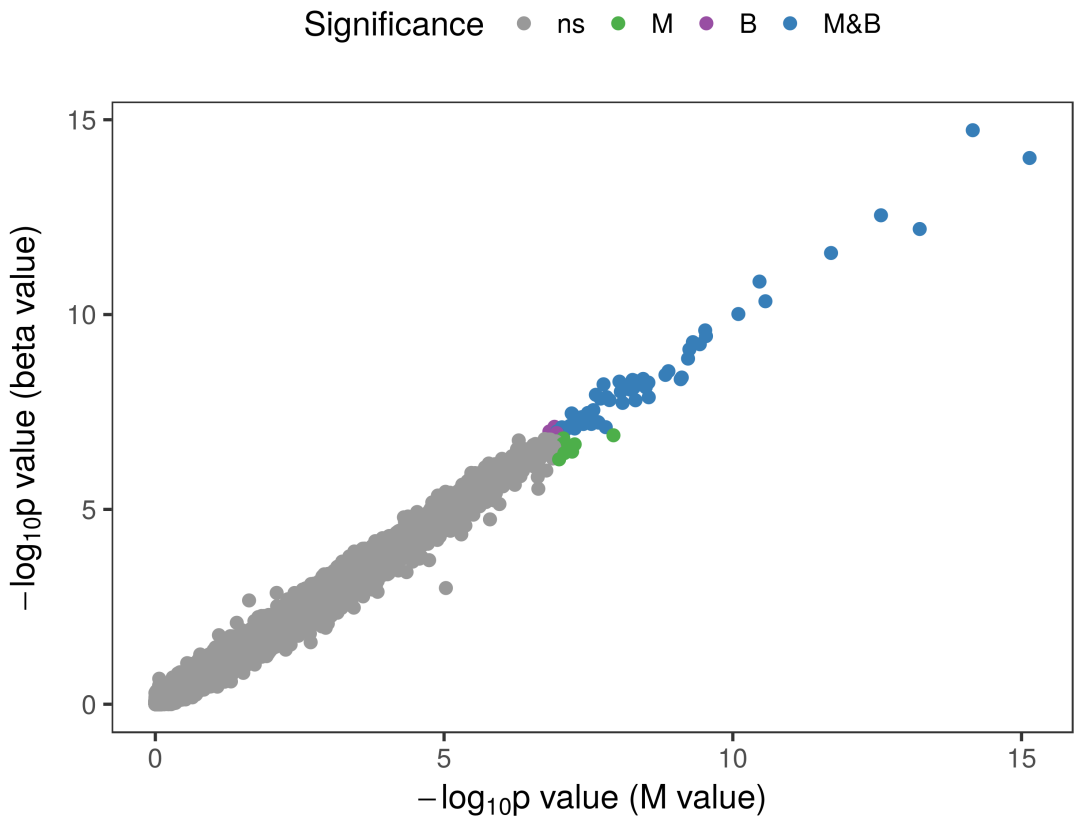


Figure 3.48: Comparison of results from analysis of the combined paediatric data using β values and M values. Points coloured by whether the probe was significant only when using M values (n=8), only using β values (n=3), or in either analysis (n=65). Spearman's ρ= 0.99, p<2.2×10^{-16}

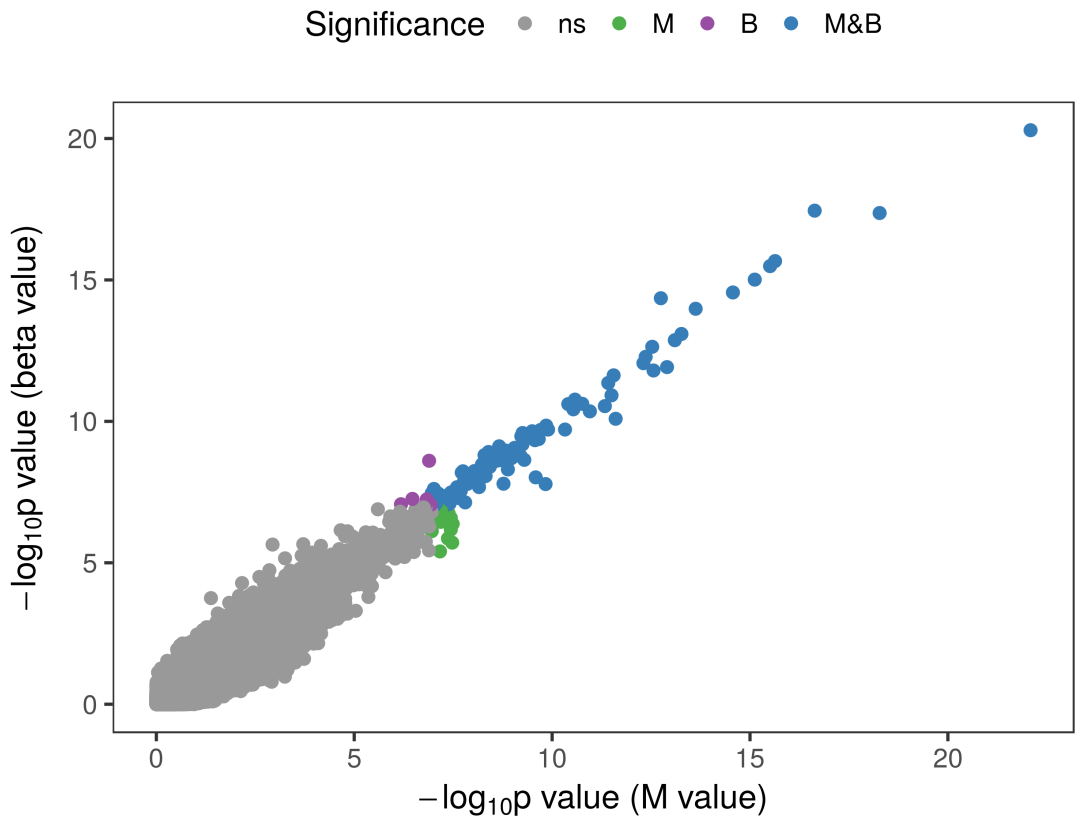


Figure 3.49: Comparison of results from analysis of the IBD Character dataset (see appendix /refch.rep) with β values and M values. Points coloured by whether the probe was significant only when using M values (n=12), only using β values (n=6), or in either analysis (n=114). Spearman's ρ= 0.92, $p<2.2\times10^{-16}$

Alternative models for analysis of methylation data

A number of alternative models have been proposed for proportion data, or methylation specifically including beta regression [184], GAMLSS[4] [519], and ratio of corrected gammas [690]. These share an approach of finding models which fit within the constraints of the data, rather than trying to reshape the data to fit the constraints of the model.

Ferrari and Cribari-Neo [184] proposed a regression method for data which are constrained to the unit scale, such as rates, probabilities, and proportions, using the beta distribution, which is itself constrained to the unit scale. The models produced deal well with heteroscedastic and skewed data, and model parameters are directly interpretable in terms of the response variable.

The beta distribution can take on a variety of shapes to fit data by varying the two parameters it takes, α and β, *e.g.* the predicted methylation profiles created by Rakyan *et al.* (and reproduced in figure 3.43, page 121). The beta distribution's probability density function (PDF) is given by:

$$f(x;\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$$

where:

$$0 \leq x \leq 1, \quad \alpha > 0, \quad \beta > 0$$

and $\Gamma(\cdot)$ is the gamma function: $\Gamma(y) = (y-1)!$ extended to all complex numbers except non-positive integers.

The mean and variance of x are related to the parameters α and β like so:

$$E(x) = \frac{\alpha}{\alpha+\beta} \quad Var(x) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

Defining $\mu = \alpha/(\alpha+\beta)$ and $\phi = \alpha+\beta$ [184] and substituting gives the mean and variance of x as:

$$E(x) = \mu \quad Var(x) = \frac{\mu(1-\mu)}{1+\phi}$$

As the variance of x is related to the mean μ, heteroscedasticity is assumed [131]. The probability density function can be redefined in terms of the mean and precision parameters μ and ϕ as:

[4]Generalised additive models for location scale and shape

$$f(x;\mu,\phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} x^{\mu\phi-1}(1-x)^{(1-\mu)\phi-1}$$

where:

$$0 < x < 1, \quad 0 < \mu < 1, \quad \phi > 0$$

The beta regression model is shown below for a random sample $(x_1, \ldots, x_n)$ of X, which is beta distributed $x_i \sim \mathbf{B}(\mu_i, \phi)$, with a vector $y = (y_1, \ldots, y_k)^T$ of k covariates, and a vector $\beta = (\beta_1, \ldots, \beta_k)^T$ of k regression parameters found by maximum likelihood estimation. Where $g(\cdot)$ is a link function which maps $(0,1) \mapsto \mathbb{R}$:

$$g(\mu_i) = y_i^T \beta$$

This model was extended to incorporate variable dispersion by Simas *et al.* [574] allowing a second set of covariates and regression parameters for the precision parameter ϕ. The covariates for μ and ϕ can overlap if desired.

An underlying assumption of beta regression is that β follows a beta distribution, implying that M and U are independent [690]; however, this is frequently untrue [353]. Weinhold *et al.* propose a model which does not assume the independence of M and U which they call the ratio of corrected gammas (RCG) [690]. Using Kibble's bivariate gamma, with the following probability density function:

$$f(x,y) = \frac{f_\alpha(x) f_\alpha(y) \Gamma(\alpha)}{1-\rho} \exp\left[-\frac{\rho(x+y)}{1-\rho}\right] (xy\rho)^{-(\alpha-1)/2} I_{\alpha-1}\left(\frac{2\sqrt{xy\rho}}{1-\rho}\right)$$

where $\Gamma(\cdot)$ is the gamma function, and:

$$x, y \geq 0, \quad \alpha > 0, \quad 0 \leq \rho < 1$$

$$f_\alpha(t) = \frac{t^{\alpha-1} e^{-t}}{\Gamma(\alpha)}$$

$$I_k(z) = \sum_{r=0}^{\infty} \frac{(z/2)^{k+2r}}{r!\Gamma(k+r+1)}$$

This is used to find the probability density function for $b = M/(M+U)$ [690]:

$$f_b(b) = \frac{\Gamma(2\alpha)}{\Gamma^2(\alpha)}(\lambda_m\lambda_u)^\alpha(1-\rho)^\alpha(b(1-b))^{\alpha-1} \times \frac{(\lambda_m b + \lambda_u(1-b))}{((\lambda_m b + \lambda_u(1-b))^2 - 4\rho\lambda_m\lambda_u b(1-b))^{\alpha+0.5}}$$

and the following log-likelihood function for methylation values and a vector of covariates $(b_1, X_1^T), \ldots, (b_n, X_n^T)$:

$$\sum_{i=1}^{n} \log(f_b(b_i, X_i; \alpha, \rho, \gamma)) = \sum_{i=1}^{n} \Big[\log(\Gamma(2\alpha)) - 2\log(\Gamma(\alpha)) + \alpha X_i^T\gamma + \alpha\log(1-\rho) + \log((\exp(X_i^T\gamma)-1)b_i + 1) + (\alpha-1)\log(b_i(1-b_i)) - (\alpha+0.5)\log\Big(((\exp(X_i^T\gamma)-1)b_i+1)^2 - 4\rho\exp(X_i^T\gamma)b_i(1-b_i)\Big)\Big]$$

These alternative techniques for modelling methylation data are clearly more computationally expensive, and it will be important to establish whether they offer a substantial improvement over currently used methods.

3.4.3 GWAS correlation

Nimmo *et al.* [458] assayed disease-associated methylation differences in 40 adult female non-smokers (21 ileal CD, 19 healthy controls) using the Illumina 27k platform. These results demonstrated an increased proportion of significant methylation differences within 25, 50 and 100kb of the 71 then-described CD GWAS risk loci [194] (figure 3.50, page 145). There was an 8.6 fold enrichment of methylation results with FDR corrected $p<0.0005$ within 50kb of the known CD GWAS SNPs ($p=0.021$). The methylation results from Illumina 450k analysis of the combined paediatric cohort were analysed to see if this finding was replicated, and similar results were obtained and presented in Adams *et al.* [4] (figure 3.32, page 98).

In this figure, methylation results with $p<10^{-10}$ were not shown because at smaller p value thresholds the list of significant probes becomes dominated by *VMP1/MIR21*, which is 47kb from the IBD GWAS SNP rs1292053 (figure 3.51, page 145). The correlation remained significant at thresholds of 45kb and 25kb, indicating that the result is not solely due to the effect of *VMP1/MIR21* methylation.

The direct comparison between the smallest p values within range of GWAS SNPs and random regions of matching probe density by Wilcoxon rank sum test is resilient to outliers. Additionally, randomly selecting control regions containing the same density of Illumina 450k methylation probes addresses the previously uncorrected bias due to the correlations between probe density and methylation results (figure 3.37, page 107), and between probe density and GWAS SNPs.

A portion of the bias is due to false positives, *i.e.* if more measurements of methylation are made in a region, there is a greater chance of a false positive result. However, the bias inherent to the design of the Illumina 450k chip is that genes which are known to be involved in disease or critical pathways, such as those implicated by GWAS studies, have better coverage by methylation probes.

The use of GWAS SNPs for other conditions, also matched for probe density, offers another potential improvement to the analysis. There was a weak, but highly significant relationship between the significance of methylation results and the significance of nearby GWAS SNPs for any trait (table 3.12, page 103), demonstrating that random regions of matched probe density are not a sufficient control for GWAS SNPs, and that GWAS SNPs for other traits should be used in similar analyses. There is some overlap between GWAS risk loci for inflammatory and autoimmune conditions, which will result in an underestimation of the true association [297, 518].

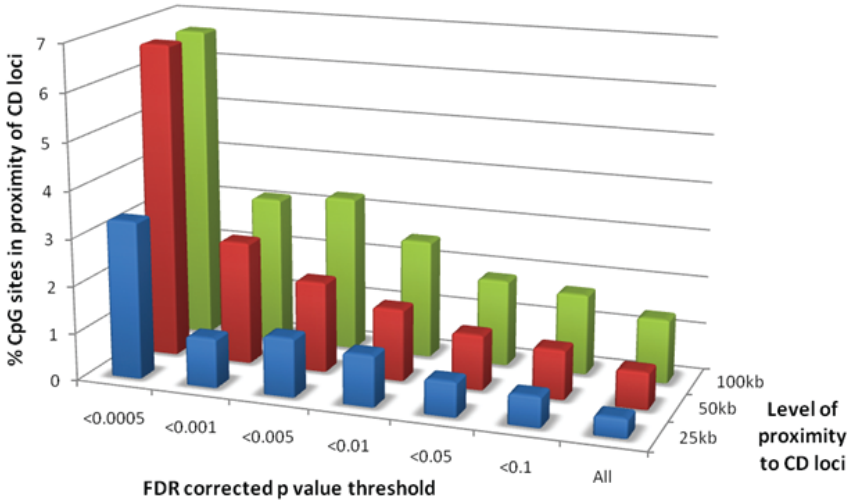


Figure 3.50: At lower p value thresholds an increasing proportion of methylation findings are within 25, 50 and 100kb of GWAS risk loci. From Nimmo *et al.* [458]

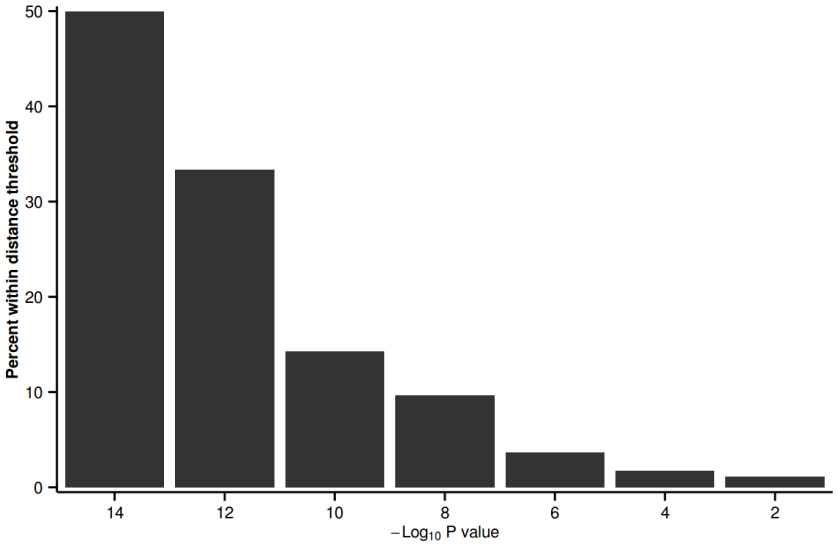


Figure 3.51: Enrichment of methylation changes within 50kb of GWAS SNPs including results with $p < 10^{-10}$

Cholesterol

In addition to CD and IBD, significant CD-associated methylation findings were significantly enriched in proximity to GWAS findings for HDL cholesterol, LDL cholesterol, and lipid metabolism phenotype (see figures 3.35 and 3.36, pages 104 and 105). Cholesterol levels are known to be deranged in IBD, with decreased total cholesterol and increased triglycerides. Differences increase in more severe disease and possibly become more normal with treatment [342].

Lower cholesterol levels were found to be associated with younger age at diagnosis, male sex, use of biologics, surgery and hospitalisations. Higher levels of triglycerides were associated with surgery, hospitalisations and median CRP [342]. Conversely, there is evidence that treatment with statins may reduce steroid requirements in IBD [132].

There are several gastrointestinal factors in IBD which would predispose to lower cholesterol levels such as malabsorption, intestinal resection, altered bowel transit, and anorexia, with possible interference from altered gut flora. Changes in lipid profile are more profound in CD compared to UC, possibly due to ileal disease interfering with bile acid and lipid absorption [342, 576]. However, similar lipid changes seen in systemic lupus erythematosus and evidence of a role for TNFα, IFNγ, and interleukins in lipid metabolism point to a non-gastrointestinal contribution [72].

Despite lower cholesterol levels, an increased risk of thrombosis and stroke is a recognized extraintestinal manifestation of IBD (table 1.2, page 6), with a two- to threefold increased risk of venous thromboembolism — an observation which has been dubbed 'the lipid paradox' [576]. Subclinical atherosclerotic changes are detectable in paediatric CD, including an increased carotid intima media thickness and reduced brachial flow-mediated dilation [10].

Chronic inflammation is an independent risk factor for atherosclerosis and thromboembolism, and is also seen in rheumatoid arthritis, systemic lupus erythematosus, ankylosing spondylitis, psoriatic arthritis, antiphospholipid syndrome, systemic sclerosis, and Sjörgen's syndrome [448]. Chronic inflammation has effects on endothelial function, immune cell activity (involved in atherosclerotic plaque formation and rupture), and levels of thrombotic and antithrombotic factors. CRP and calprotectin are frequently elevated in IBD, and beside use as inflammatory markers may be actively involved in atherogenesis. Other products which are increased in IBD have been linked to atherosclerosis, such as antiphospholipid antibodies, homocysteine and lipoprotein [576, 448].

NOD2 mutations also predispose to premature atherosclerosis and destabilisation of coronary plaques and IBD patients have thrombocythaemia, with small platelets containing increased granules and demonstrating a propensity for spontaneous aggregation [576].

Glucocorticoids increase hypertension, weight gain, poor glucose control, dyslipidaemia and cardiovascular events, though in inflammatory conditions the positive effects on inflammation may partly or wholly offset this. Additionally anti-TNFα drugs, methotrexate and 6-mercaptopurine may also have beneficial effects on atherosclerotic risk [448].

The relationship between cholesterol, atherosclerosis, and IBD is clearly complex, and incompletely understood, but consistent with the finding of methylation changes enriched in proximity to GWAS loci with established links to HDL cholesterol, LDL cholesterol and lipid metabolism. Aloi *et al.* described subclinical atherosclerotic findings in children with IBD [10] and if these methylation findings support an ongoing atherosclerotic process in these children it would lend weight to investigating very early interventions to reduce lifetime risk for CHD, VTE, and stroke.

3.4.4 Transcription factors

Stadler *et al.* described low-methylated regions (LMRs) in murine embryonic stem cells and neural progenitors and human stem cells [588]. These LMRs were found distal to promoters and TSS, were mostly outwith CGIs and had low CpG density. They associated with DNA-binding factors, contained DNase hypersensitivity sites, and histone marks associated with active enhancers such as H3K4me1 and H3K27ac, but with little RNA polymerase II activity indicating they were unlikely to act as promoters. Transcription factor binding was shown to cause locally reduced methylation, LMR formation was strongly correlated with increased expression of the nearest gene ($p=2.3\times10^{-197}$) and LMRs were shown to be cell type specific. LMRs have also been shown in differentiated cells such as B cells and to correlate with numerous DNA-binding factor motifs [80].

For the transcription factor CTCF, DNA methylation surrounding binding sites was shown to be decreased in CTCF-bound DNA segments regardless of whether the binding site itself contained a CpG, and it was found that CTCF bound to both methylated and unmethylated regions but with an inverse correlation between methylation and binding strength [179]. Decreased methylation and increased hydroxymethylation subsequent to transcription factor binding suggests an induction of DNA methylation turnover [179].

LMRs were described as having intermediate methylation (mean 30%, but highly variable) [588, 179], and as shown in section 3.3.2 the strongest methylation findings in this study demonstrate intermediate methylation levels (figure 3.4, page 64) which is a significantly different ($p<2.2\times10^{-308}$) pattern to other probes which are almost all nearly fully methylated or unmethylated. The most significant probes predominantly showed hypomethylation in CD (76.9% vs 43.3% for all probes, χ^2 $p=9.03\times10^{-8}$), and section 3.3.3 showed that Bonferroni and FDR significant probes were significantly less likely to be in CGIs compared to random non-significant probes — 31.1% (χ^2 $p=4.58\times10^{-6}$) and 19.4% ($p=1.06\times10^{-166}$) versus 60.7% (95% CI 60.0–61.4%).

It would therefore be consistent for a portion of the highly significant methylation differences described in this study to represent increased occupancy of enhancers or other distal regulatory regions by DNA-binding factors. Further work would be required to confirm this, such as CHIP-seq or assaying 5hmC.

The DNA motif `CTDGNGGT` was identified as being enriched among Bonferroni (35.1%, p=0.038) and FDR significant (23.2%, $p=2.8\times10^{-5}$) probes (section 3.3.7). As shown in table 3.18 (page 115), 30% of these probes had no nearby

transcripts, and none were in a clear promoter site, though 20% could be in alternate promoters for shorter transcripts, or in the case of *VMP1*, a microRNA. However, there were indications of an active regulatory region at 90% of the motifs, with ENCODE finding transcription factor binding sites by ChIP-seq, H3K27 acetylation and DNaseI hypersensitivity. This motif was identified as sharing high homology with binding sites for two human transcription factors — GLI2 and RUNX3 (table 3.16, page 112).

GLI

GLI2 is a glioma-associated oncogene and zinc-finger transcription factor, also associated with holoprosencephaly, pituitary abnormalities, cleft lip and polydactyly [56, 121]. The GLI family have a shared binding site, overlapping expression and function, but some important differences in regulation and targets [513, 533, 311]. There is also intrafamily regulation, with GLI2 activating GLI1 [147].

GLI proteins are critically important to the Hedgehog signalling pathway [311], which is involved in embryonic development, bone growth, stem cell maintenance, cell cycle regulation, apoptosis, and cancer [147, 513]. Hedgehog signalling is also involved in inflammation leading to increased expression of IL-6, TNFα, IL-12, CCl5, and CXCL9 [210]. Non-Hedgehog-related consequences of GLI mutations, and expression during embryonic development in regions without *Shh* production imply that GLI proteins have some Hedgehog-independent actions [147]. GLI1 and GLI2 have been shown to act independently of Hedgehog signalling, and are induced by TGFβ in a SMAD3 dependent manner [147].

The *GLI1* gene is within the IBD2 risk locus (12q13) and sequencing revealed a nonsynonymous SNP reducing the activity of GLI1 by half [366]. This SNP was associated with IBD, particularly UC, with a meta-analysis of 5352 patients finding an odds ratio of 1.19 [366]. $GLI1^{+/\mathrm{lacZ}}$ mice were more severely affected by DSS colitis, with increased expression of pro-inflammatory cytokines IL-23, IL-12 and IL-17 [366]. NOD2 activation leads to iNOS expression and subsequently SHH signalling via NFϰB and NOD2-induced mir146a reducing levels of NUMB – a negative regulator of GLI [210].

RUNX3

The Runt-related transcription factors RUNX1, RUNX2, and RUNX3 have high sequence homology and a shared binding site. RUNX proteins form a heterodimer with CBFβ [69] and bind specific DNA sequences through the terminal 128aa conserved from *Drosophila* transcription factor *runt* — the *runt* homology domain (RHD) [475]. The binding site is commonly cited as `5′-RACCRCA-3′`, `5′-TGYGGT-3′`, or `5′-AACCACA-3′`.

The RUNX proteins are exchangeable *in vitro*, but have different expression profiles and regulation *in vivo* [475]. In mice *RUNX1* KO prevents mature haematopoietic differentiation [475], *RUNX2* KO mice have no bone development [475], and loss of *RUNX3* causes spontaneous colitis and gastric mucosal hyperplasia [76].

All *RUNX* genes are regulated by retinoids and vitamin D [475]. RUNX3 is involved in T cell and proliferation and differentiation into $CD8^+$ T cells by Th-POK suppression [562] and together with T-bet into T_H1 [728], with critical functions and high expression in both lineages [321, 154, 608]. RUNX3 is targeted by TGFβ1 [689] and is involved in TGFβ signalling and the maturation of dendritic cells and activation of macrophages [689]. RUNX3 also functions as a tumour suppressor and apoptosis inducer, and has a role in neural development [287].

A potential role for RUNX3 has been proposed in several inflammatory diseases including IBD, RA, SLE, ankylosing spondylitis and psoriasis [689]. *RUNX3* is in an IBD risk locus (IBD7, 1p36) with possible epistatic interactions with *NOD2* [115, 116, 711]. Weersma *et al.* identified a significant association with UC for *RUNX3* SNP rs2236851, and epistatic interactions between *RUNX3* and *SLC22A4/5* polymorphisms were implicated [689]. One study showed no association between 11 examined *RUNX3* SNPs and IBD (10 promoter and one intronic SNPs), however the authors acknowledge that the SNP from Weersma *et al.* was not tested and was not in LD with the tested SNPs [229].

RUNX3 is also in a GWAS locus for ankylosing spondylitis [669] — a chronic inflammatory condition of the axial skeleton which is frequently comorbid with IBD (5–10%) [661], shares family histories [619] and genetic risk factors [518]. *SLC22A4* and *SLC22A5* are associated with CD (IBD5) and a polymorphism in *SLC22A4* disrupting the RUNX binding site has been linked to rheumatoid arthritis [626]. *RUNX3* has been reported as down-regulated in CD [248] and overexpressed in inflamed colonic mucosa [689], with IHC showing nuclear stain-

ing in sub-epithelial T-lymphocytes in small and large intestine [689].

RUNX3 hypermethylation has been found in UC-associated CRC (UC-CRC) patients compared to UC patients and healthy controls, both in the tumour and in distant biopsies [203], and in UC versus controls [231]. *RUNX3* inactivation is implicated in multiple types of cancer, more frequently due to promoter methylation than mutation [644, 230], and has been proposed as a marker of CpG island methylator phenotype (CIMP) type tumours [433]. A methyl donor enriched diet in mice caused transgenerational *RUNX3* hypermethylation and increased response to an allergic disease model [267].

Each *RUNX* gene is transcribed from two promoters, each containing RUNX binding sites [475, 69, 321]. The RUNX3 protein derived from the distal promoter is highly expressed in $CD8^+$ and T_H1 cells and exhibits a greater ability to induce transcription due to a novel activation domain in the 19 N-terminal amino acids [118, 321]. The proximal product is expressed at lower levels in developing thymocytes, naive $CD4^+$, and T_H2 cells [118, 321].

RUNX3 contains a Bonferroni significant CD-associated hypermethylation ($p=2.45\times10^{-8}$, table 3.9, page 73) within a DMR (table 3.10, page 75) at the start of the distal transcript (see figure 3.18, page 84), and the associated GLI/RUNX3 motif identified in section 3.3.7 is a confirmed RUNX site. Additionally, analysis of a microarray study on intestinal biopsies previously performed by our group [462] shows increased expression in inflamed biopsies from CD patients (figure 3.52).

RUNX3 interacts with known IBD associated mechanisms including NOD2 interactor [459] TLE to epigenetically silence *CD4* in $CD8^+$ cells [719], CD-associated genes *SLC22A4/5* [689], IBD associated [77, 652] MDR1 [230] and increases expression of mir30a which directly targets the 3′ UTR of CD-associated gene vimentin (page 16 and Stevens *et al.* [590]) leading to a decrease in expression [386]. Additionally RUNX3 causes expression of IFNγ, eomesodermin (EOMES), granzyme B (GZMB) [321], and 15 microRNAs [386], suppresses expression of IL-4 [321], *MRP1*, and core apoptosis gene *BCL2* [230].

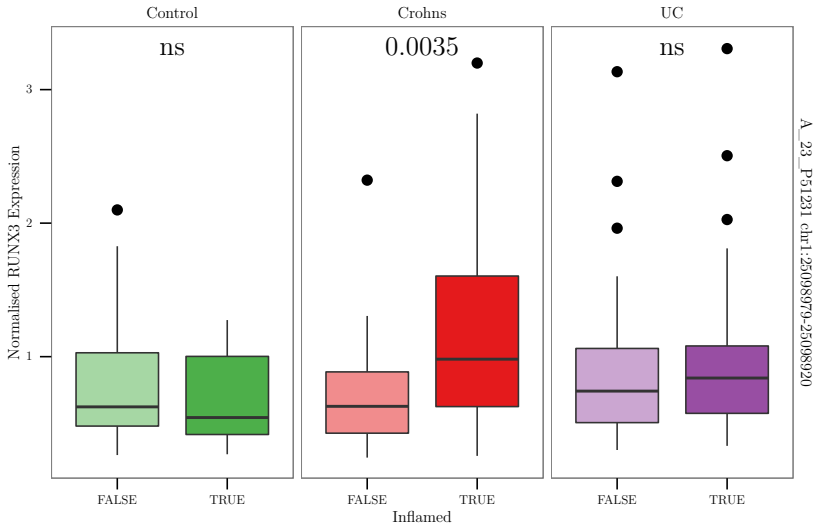


Figure 3.52: Box and whisker plot showing *RUNX3* expression in inflamed and non-inflamed biopsies from controls (56 vs. 17), Crohn's disease (48 vs. 51), and ulcerative colitis (65 vs. 64) [462]. Whiskers extend up to 1.5 × IQR from the 1st and 3rd quartiles, with any remaining results plotted as points. P values for inflamed vs non-inflamed by Wilcoxon rank sum test shown for each diagnosis. Reanalysis of data from two papers by Noble *et al.* [461, 462], see section 3.2.1, page 52 for further details.

3.4.5 Genes of interest

This section contains a brief description of genes drawn from both individual CpGs and DMRs, which are of interest to IBD. Those genes which are discussed at length in other sections are only briefly mentioned here, and genes which are part of our groups ongoing research are examined in more depth.

VMP1 The DMR within *VMP1* is the most significant, containing the most significant individual CpG, and is entirely composed of Bonferroni significant CpGs, it is also located within a GWAS risk locus for IBD. VMP1 is an essential component of the autophagy pathway, and the position of the DMR around the 11^{th} exon corresponds to the primary transcript of microRNA mir21, which has been implicated in various cancers and inflammatory conditions, including IBD. VMP1 and mir21 are discussed at greater length in chapter 4.

RPS6KA2 RPS6KA2 contains the second most significant individual CpG, it is a kinase involved in the MAPK/p38 pathway found in an IBD GWAS risk locus, and is discussed at length in chapter 5.

RUNX3 The RUNX3 DMR is located within the 5′ UTR of the longer transcript, and contains a Bonferroni significant probe ($p=2.19\times10^{-7}$). The RUNX3 gene is located within an IBD risk locus, loss of RUNX3 function in mice causes spontaneous colitis and gastric mucosal hyperplasia [76], analysis of putative DNA binding sites in proximity to methylation results (section 3.3.7, page 111) shows an enrichment of the RUNX3 binding site. RUNX3 function and findings relating to IBD are discussed further in section 3.4.4 (page 150).

TNF Tumour necrosis factor (TNFα) is a critical inflammatory cytokine, which is central to many pathways involved in IBD pathogenesis as discussed in chapter 2, and inhibition of TNFα by monoclonal antibody drugs such as infliximab is a highly valuable treatment for IBD and other inflammatory conditions such as rheumatoid arthritis, psoriasis, and ankylosing spondylitis. The *TNF* gene is located in one of the strongest genetic risk loci for IBD, and many products which interact with or modulate TNF have also been identified as being in genetic risk loci [297].

SOCS3 Suppressor of cytokine signalling (SOCS)[5] proteins act a negative feedback and regulatory mechanism to cytokine signalling [722]. The SOCS box domain of SOCS family proteins combines with elongin B and C to function as a ubiquitin E3 ligase which marks bound substrates for proteosomal degradation [320, 197, 382]. SOCS protein specificity is determined by the central SH2 domain which interacts with the cytoplasmic portions of gp130-related cytokine receptors after specific tyrosine residues have been phosphorylated [320, 722],

The JAK-STAT pathway is activated by multiple cytokines including interleukins, interferons and haematopoietic growth factors [722]. SOCS3 acts as a negative regulator of the JAK-STAT signalling pathway by interacting with STAT3 [530], the gene for which is in a gwas risk locus for IBD [297] and by direction inhibition of JAK tyrosine kinase activity through the kinase inhibitory region [722]. Many other JAK-STAT genes can be found in IBD risk loci including *JAK2*, *SOCS1*, *STAT1* and *STAT4* [297].

SOCS3 is important in NOD2 tolerance to MDP. SOCS3 expression is induced by MDP in an NOD2 and RIPK2 dependent manner. SOCS3 then binds to the CARD domain of activated NOD2 and causes proteosomal degradation of NOD2 and RIPK2 [363].

SOCS3 also regulates other pathways such as IFNβ signalling in the innate antiviral response [382] and IL-1 induced NFκB activation by TRAF6/TAK1 signalling [197].

SOCS3 is induced by IL-6/STAT3 [737, 135], all three of which are increased in inactive and active UC biopsies compared to healthy controls [376] and acts to shorten and regulate the response to IL-6 [135, 356]. Elimination of SOCS3 binding allows IL6R to generate an IL10-like anti-inflammatory response [170]. Loss of SOCS3 in mice increased the oncogenicity of DSS and TNFα mediated NFκB activation [520]. SOCS3 expression in known to be regulated by DNA methylation and microRNAs in various cancers [71] and *SOCS3* promoter methylation and loss of expression was found in the progression to UC-associated colorectal cancer [376]. MicroRNA mir19b directly downregulates *SOCS3* expression levels and is underexpressed in Crohn's disease, and delivering mir19b to the colon protected against TNBS-induced colitis in mice [112].

Silencing mir122 is a therapeutic strategy in hepatitis C (HCV) and is the first use of an antimir drug (Miravirsen) in humans [293]. Silencing mir122 has been found to significantly increase methylation at the *SOCS3* promoter in a DNMT1

[5]Also called SSI (STAT-induced STAT inhibitor)

independent mechanism, decreasing SOCS3 expression and enhancing STAT3 activation [721], it is also notable that mir122 directly inhibits *NOD2* expression [111] (see page 34). Increased hepatic expression of SOCS3 is correlated with a poor response to treatment in HCV and HIV and is a stronger predictor than viral genotype, and higher levels of SOCS3 in African-Americans could explain a worse response to antiviral therapy in African-Americans [322]. The reduction in *SOCS3* by antimir122 causing an increase in IFNα signalling could be partially responsible for the antiviral effects of this compound [721] in addition to abolishing the direct stabilising effect mir122 has on the HCV genome [293].

Inflammatory cytokines such as TNFα and IL-6 are important promoters of tissue repair, and it has been proposed that the increased expression of SOCS3 seen in IBD may limit epithelial repair, perpetuating an inflammatory state despite its anti-inflammatory actions [614].

Evidence from mice shows that an appropriate response to cytokines depends on the proper balance of SOCS proteins, and that deficiency in one may lead to the activation of alternative responses; for example SOCS3 prevents an IFNγ-STAT1 expression pattern in response to IL-6 [135, 356]. There are also interactions between the different SOCS and STAT proteins such as SOCS2 degrading SOCS3 [609]. IL6 antagonist drugs such as tocilizumab[6] and sirukumab[7] have recently become available, with more in development, and are in use for rheumatoid arthritis and systemic juvenile idiopathic arthritis. Anti-IL6 has also been explored as a treatment for CD [291].

In addition to IL-6 and MDP [363], there are many reported inducers of SOCS3 including TNFα, LPS [68], NFκB [197], IFNγ[135, 482], IL-2 [122], IL-10 [171] and TSH [482]. Other reported SOCS3 targets include IGF1 receptor [149], insulin receptor [149], epidermal growth factor receptor [707], JAK1 [122], STAT1 [135] and growth hormone receptor [510].

Polymorphisms in *SOCS3* have been linked to infantile asthma, Graves' opthalmopathy, chronic hepatitis C, atopic dermatitis, insulin resistance and obesity [175, 712]. A recent EWAS study of type 2 diabetes in a Qatari population also found significant hypomethylation of the same *SOCS3* Illumina 450k probe, together with *SBNO2* [6].

[6] Anti-IL6R
[7] Anti-IL6

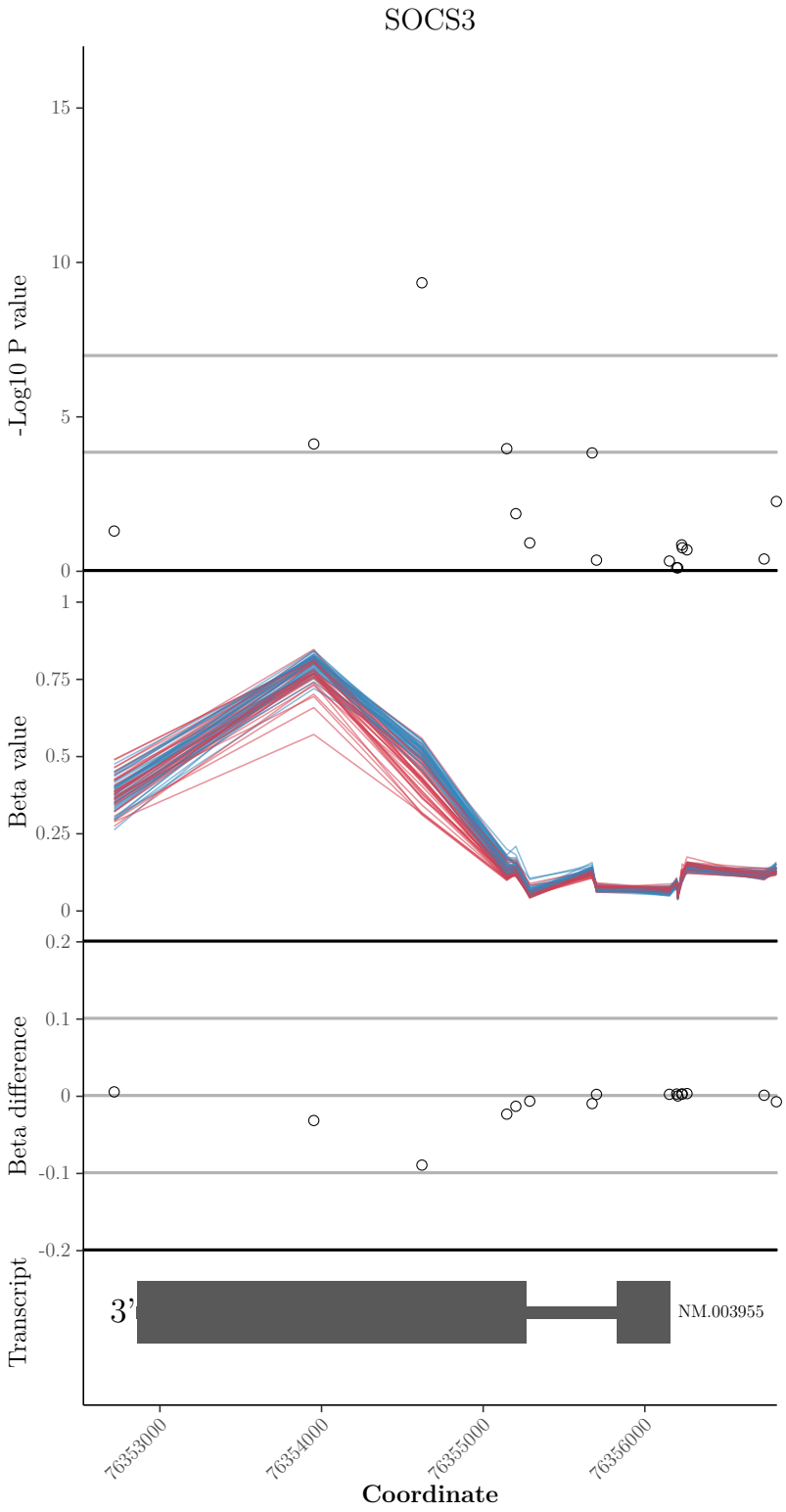


Figure 3.53: Methylation results for the SOCS3 region

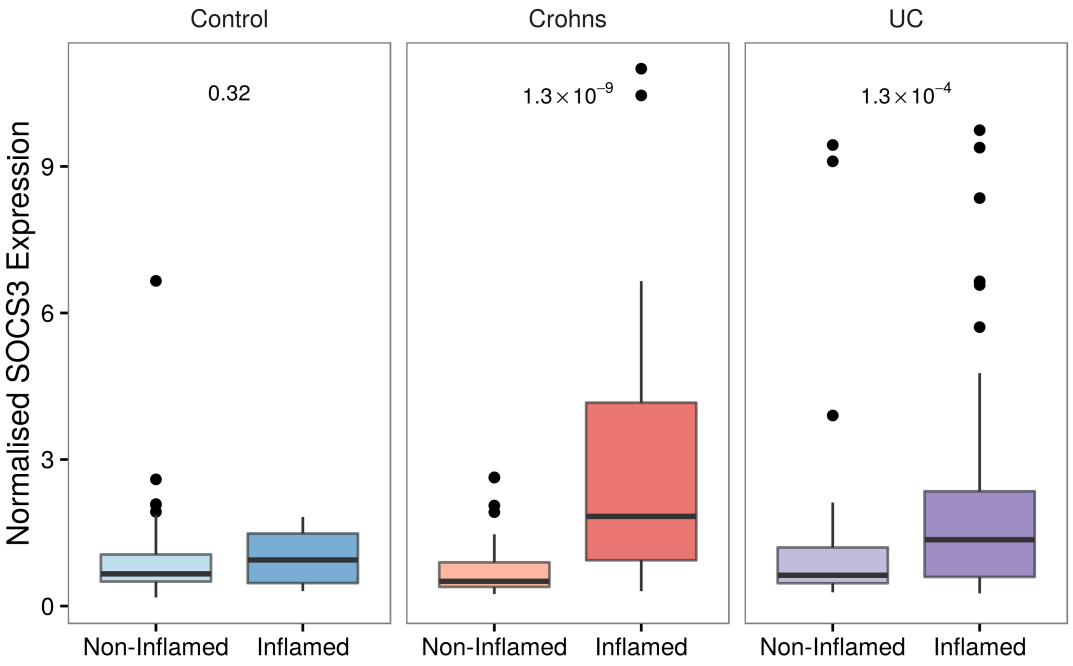


Figure 3.54: Box and whisker plot showing microarray data for *SOCS3* expression in non-inflamed and inflamed biopsies from controls (56 vs. 17), Crohn's disease (48 vs. 51), and ulcerative colitis (65 vs. 64). Whiskers extend up to 1.5 × IQR from the 1st and 3rd quartiles, with any remaining results plotted as points. P values for inflamed vs non-inflamed by Wilcoxon rank sum test within each diagnosis. Reanalysis of data from two papers by Noble *et al.* [461, 462], see section 3.2.1, page 52 for further details.

SBNO2 *SBNO2* is located within a CD GWAS risk locus [297, 194, 41] with negative [17] or weak [643] findings in UC. Although this finding has been well replicated there is very little published research on this gene in humans. Increased expression was found in a B-cell derived cell line in response to very high fluoride exposure [253], there was one report of a link to schizophrenia [364], and a role in osteoclast fusion [413].

SBNO2 is a homologue of the *Drosophila* gene *strawberry notch* (*sno*), a downstream positive regulator of the *Notch* signalling pathway expressed in the nucleus [404]. Activation of the Notch receptor causes the release of the Notch intracellular domain which migrates to the nucleus and regulates transcription of downstream genes. The Notch signalling pathway is evolutionarily conserved and is involved in development, differentiation, barrier function, immune response and apoptosis [416, 465]. Notch knockdown worsened colonic response to DSS, dampened the chemokine response to colonic injury [416]

SBNO2 is is a potent inhibitor of NFκB-mediated transcription [171], and is induced by IL-6, IL-10, IL-1β, LPS and TNFα [225, 279, 171] and in *E. coli* sepsis in mice [61]. SBNO2 has a level of background expression which is STAT3 independent, but IL-10 upregulation is STAT3 dependent [171].

Transcription factor STAT3 binds to SBNO2 resulting in increased expression and a change in transcript [279] and there is evidence of transcript switching after IL-6 signalling in mouse astrocytes [225]. There are two recognized *SBNO2* transcripts in humans, with one having a shortened N-terminal end. Zebrafish have 3 *sno* genes; *sbno1*, *sbno2a* (chromosome 2) and *sbno2b* (chromosome 22) [603], with *sbno1* and *sbno2a* showing similar expression profiles and functions and *sbno2b* exhibiting a different pattern and a role in neurodevelopment [602].

In humans, *SBNO1* has been identified as a GWAS risk locus for lipid levels [613] height [357], and schizophrenia [522] and infant head circumference [67] possibly indicating a preserved role in neurodevelopment. Despite high homology and related functions SBNO1 does not compensate for a loss of SBNO2 [171].

Mutations in *STK11*, adjacent to *SBNO2*, cause Peutz-Jeghers syndrome (PJS) — an autosomal dominant condition causing hyperpigmented mucocutanous lesions and generally benign hamartomatous gastrointestinal polyps, but with a high life-time cancer risk. In 30% of cases there is a large deletion which can affect multiple genes including *SBNO2*, however, haploinsufficiency for *SBNO2* in these patients did not alter the PJS phenotype [515].

IL-10 has a critical role in limiting inflammation [708]. *IL10*$^{-/-}$ mice were found to develop spontaneous enterocolitis, and their CD4^{+} cells caused colitis

when transplanted into immunodeficient mice [514]. The colitis appeared to be due to uncontested IL-12 and IFNγ signalling and overproduction of T_H1 cells, and could be prevented by anti IL-12 or anti-IFNγ. As a side effect of decreasing IFNγ production, IL-10 does however also retard bacterial clearance [514]. IL-10 deficient mice are healthy in germ-free conditions, but develop inflammation in response to normal bowel flora [561].

IL-10 reduces inflammatory mediators (IL-12, TNFα, IL-1α, IL-1β, and IL-6 [170]), inhibits antigen presentation and promotes epithelial integrity. JAK1-STAT3 signalling is essential for IL-10 induced anti-inflammatory effects [170, 450]. The main sources of IL-10 are monocytes, macrophages, dendritic cells and multiple T_H subsets (including T_H1, T_H2, T_H17, and all T_{reg} subtypes) [538, 450].

There is significant evidence for the involvement of IL-10 in IBD. *IL10* is in an IBD GWAS locus [297, 191] and IL-10 production is impaired in severe IBD [127, 410]. The CD-associated *NOD2* mutations *3020insC*, R702W, and G908R actively inhibit *IL10* transcription by reducing p38-mediated phosphorylation of IL-10 promoting transcription factor hnRNP-A1 [463].

Glocker *et al.* described *IL10RA* and *IL10RB* mutations in cases of very severe enterocolitis occurring in the first year of life. These patients had a reduced STAT3 response to IL-10 and increased production of TNFα and other inflammatory cytokines. These patients also showed a lack of SOCS3 up-regulation in response to IL-10 as opposed to their unaffected siblings [213]. Further studies have replicated the findings of *IL10* and *IL10RA/B* variants being strongly correlated with onset of IBD in the first year of life, a positive family history and sometimes consanguinity, a severe clinical course, and poor response to conventional treatment but positive response to autologous haematopoietic stem cell transplantation [567, 566, 57, 339].

Several studies have examined IL-10 as a therapeutic agent in Crohn's disease. A Cochrane review in 2010 [85] found no statistical difference from placebo - although the authors only considered three studies [650, 178, 557]. A review of the literature including animal studies found a mild benefit to IL-10 treatment, but overall the results were disappointing [410]. It was already known that IL-10 is effective in preventing inflammation, but once a T_H1-mediated inflammatory state is established, it is insufficient to reverse the inflammation [514]. There is evidence that delivery of IL-10 to the intestinal lumen by *Lactococcus lactis* engineered to secrete IL-10, adenoviral vectors, or gelatine microspheres containing IL-10 is more effective than systemic administration. Finally, some IBD patients have been shown to have normal or elevated levels of IL-10, and the treatment may

only be effective in those with deficient IL-10 production [410], as seen in the extreme phenotype of children with *IL10* or *IL10R* mutations and a subset of adult cases.

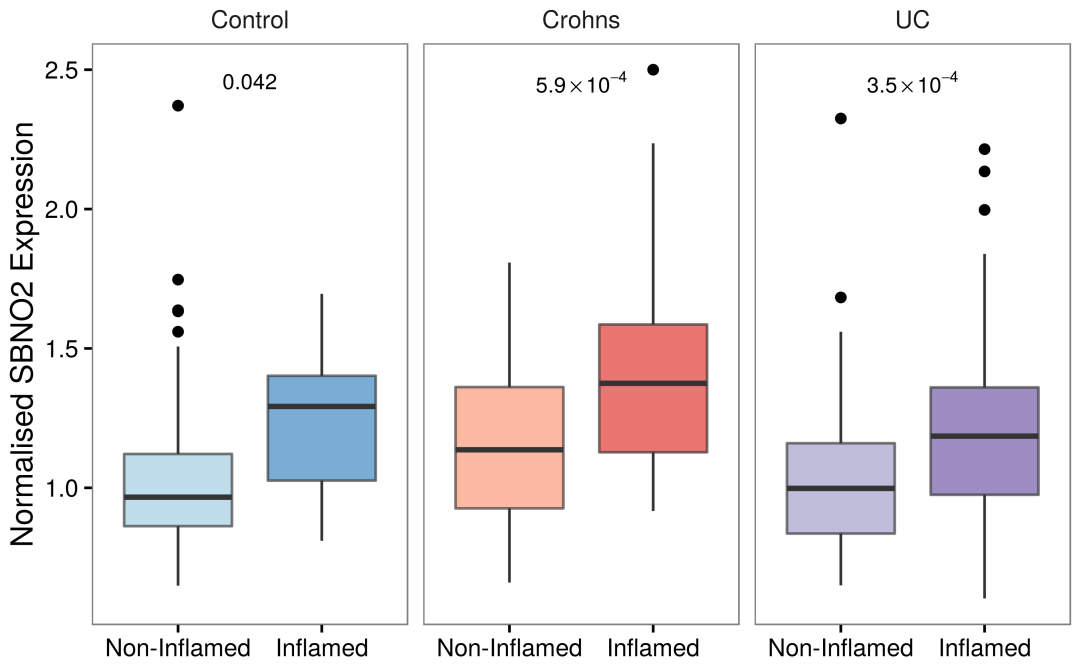


Figure 3.55: Box and whisker plot showing microarray data for *SBNO2* expression in non-inflamed and inflamed biopsies from controls (56 vs. 17), Crohn's disease (48 vs. 51), and ulcerative colitis (65 vs. 64). Whiskers extend up to 1.5 × IQR from the 1st and 3rd quartiles, with any remaining results plotted as points. Reanalysis of data from two papers by Noble *et al.* [461, 462], see section 3.2.1, page 52 for further details.

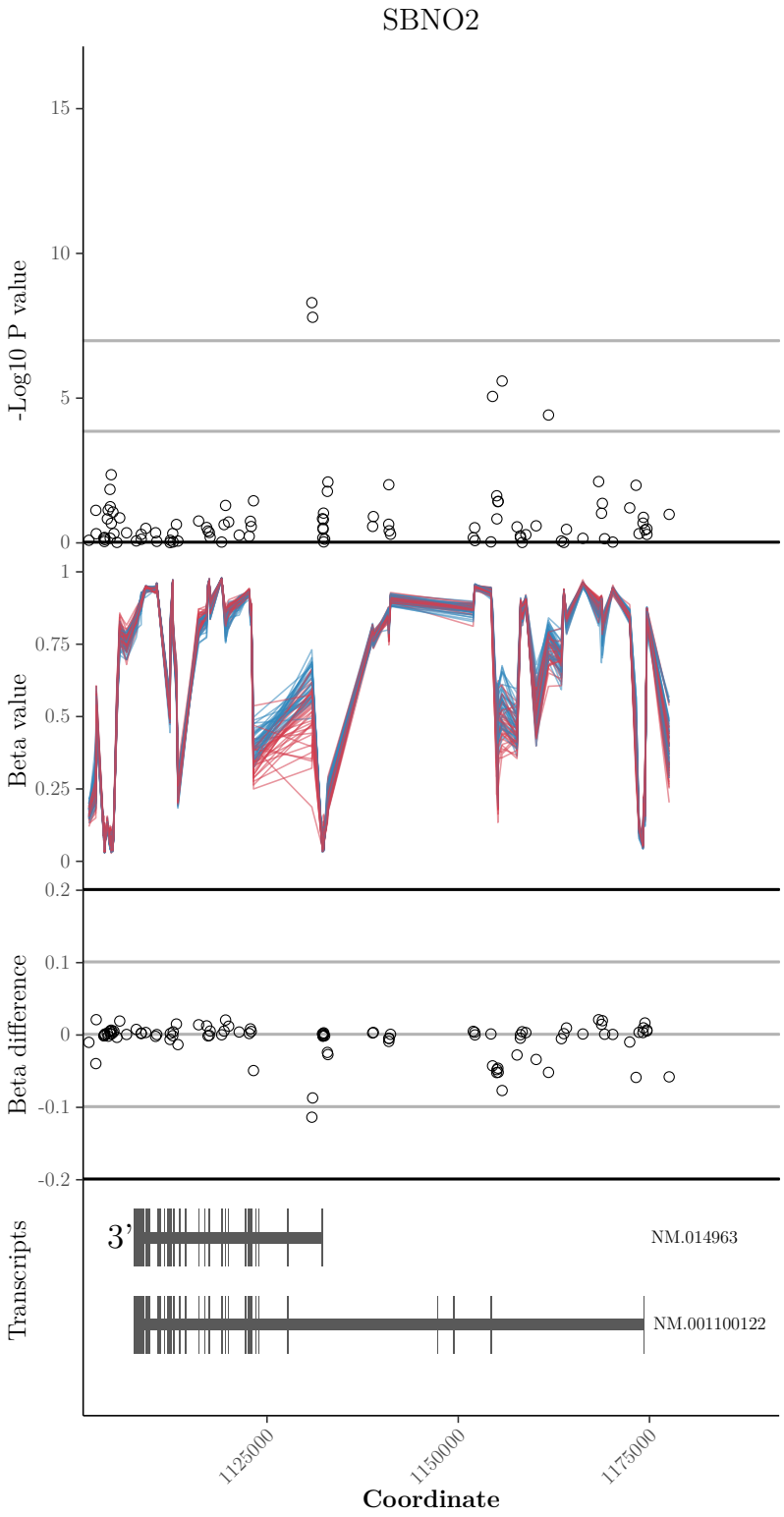


Figure 3.56: Methylation results for the SBNO2 region

NLRC5 NLRC5 is in the same subtype of the NLR family as NOD2 (CARD containing NLRC, NOD2 is NLRC2). Also known as NOD27 [624], NLRC5 is the largest of the NLR family [355] and is most closely related to NLRC1 (NOD1), NLRC2 (NOD2), NLRC3 (NOD3), and NLRA (CIITA) [454, 444]. NOD2 is the strongest genetic risk factor for Crohn's disease is discussed in section 2.2.1 (page 11). Several other NLR family members have been associated with chronic inflammatory diseases [348, 623], SNPs in the region of NLRP3 have been linked to CD [668] and NOD1 SNPs show a correlation with IBD in under-40s [395].

NLRC5 is a well-conserved cytosolic and nuclear [53, 426] protein expressed in immune tissues (bone marrow, lymph nodes, thymus, and spleen), organs with mucosal surfaces (small intestine, colon, lung, and uterus) [355, 53], and immune cells [454, 53].

Like other NLRCs, NLRC5 contains an N terminal CARD domain (though only one as opposed to two in NOD2, and with an atypical [444, 233], highly conserved [53] structure), a central NACHT-NOD domain and a LRR domain. The LRR is unusual in that it is much longer than typical [348, 444] containing 33–43 LRRs as opposed to 9–10 in NOD1 and NOD2, though they are not irregularly positioned [444] as previously reported [348].

There are 5 cDNA splice variants differing in LRR length and tissue distribution [454] and possible translational modifications [348] both of which could contribute to the conflicting results observed in functional studies.

NLRC5 expression is strongly induced by IFNγ via JAK/STAT1 and NFκB [348, 53, 426]. Response to other triggers such as LPS, CMV[8], and poly(I:C)[9] is inconsistent and may be cell type specific [348, 136, 53].

NLRC5 is a potent negative regulator of NFκB which prevents activation by multiple signals including TNFα, LPS, IL-1β, and MyD88 [136, 53]. One proposed mechanism is that NLRC5 competes with NEMO to form a complex with IKKα and IKKβ, (dependant on the LRR domain) and inhibits their kinase activity and their own phosphorylation [136].

The most closely related NLR gene to NLRC5 is CIITA (NLRA), is well established to play a vital role in regulating MHC class II genes [444, 427, 426]. IFNγ-induced NLRC5 has been shown to trigger the expression of MHC class I, and antigen processing and presentation genes including HLA-A, –B, –C, –E, β2M, LMP2, and TAP1 by interacting with promoters regions [426, 427].

NLRC5 has also been proposed to be a negative regulator of type I inter-

[8]Cytomegalovirus

[9]polyinosinic:polycytidylic acid - a TLR3 activator with similar structure to dsRNA

feron signalling in the antiviral immune response by inhibiting IFNβ signalling by blocking binding of MAVS to RIG-1 and IRF3 phosphorylation [136, 454, 233], though this could not be confirmed in NLRC5-deficient mice [355].

The methylation results in the NLRC5 region are shown in figure 3.58 (page 164). This has been extended to show the DMR in an intron of the adjacent gene CETP[10] (highlighted in grey). Analysis of previously published expression data from intestinal biopsies shows increased *NLRC5* expression in response to inflammation in CD, UC, and controls (figure 3.57, page 163).

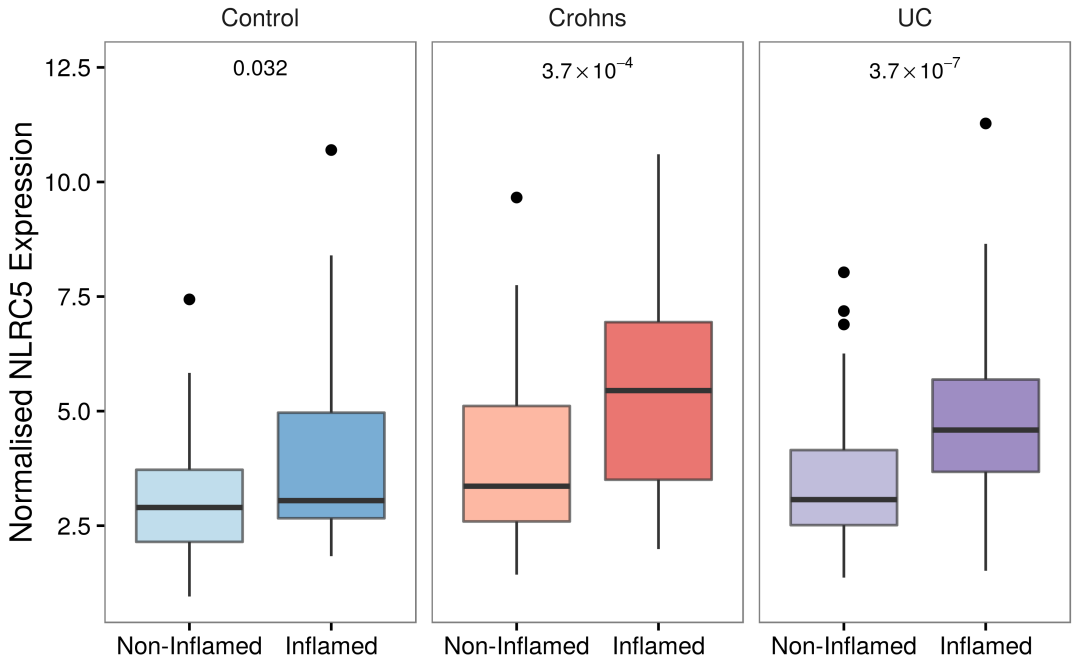


Figure 3.57: Box and whisker plot showing microarray data for *NLRC5* expression in non-inflamed and inflamed biopsies from controls (56 vs. 17), Crohn's disease (48 vs. 51), and ulcerative colitis (65 vs. 64). Whiskers extend up to 1.5 × IQR from the 1st and 3rd quartiles, with any remaining results plotted as points. P values for inflamed vs non-inflamed by Wilcoxon rank sum test within each diagnosis. Reanalysis of data from two papers by Noble *et al.* [461, 462], see section 3.2.1, page 52 for further details.

[10] table 3.10, page 75

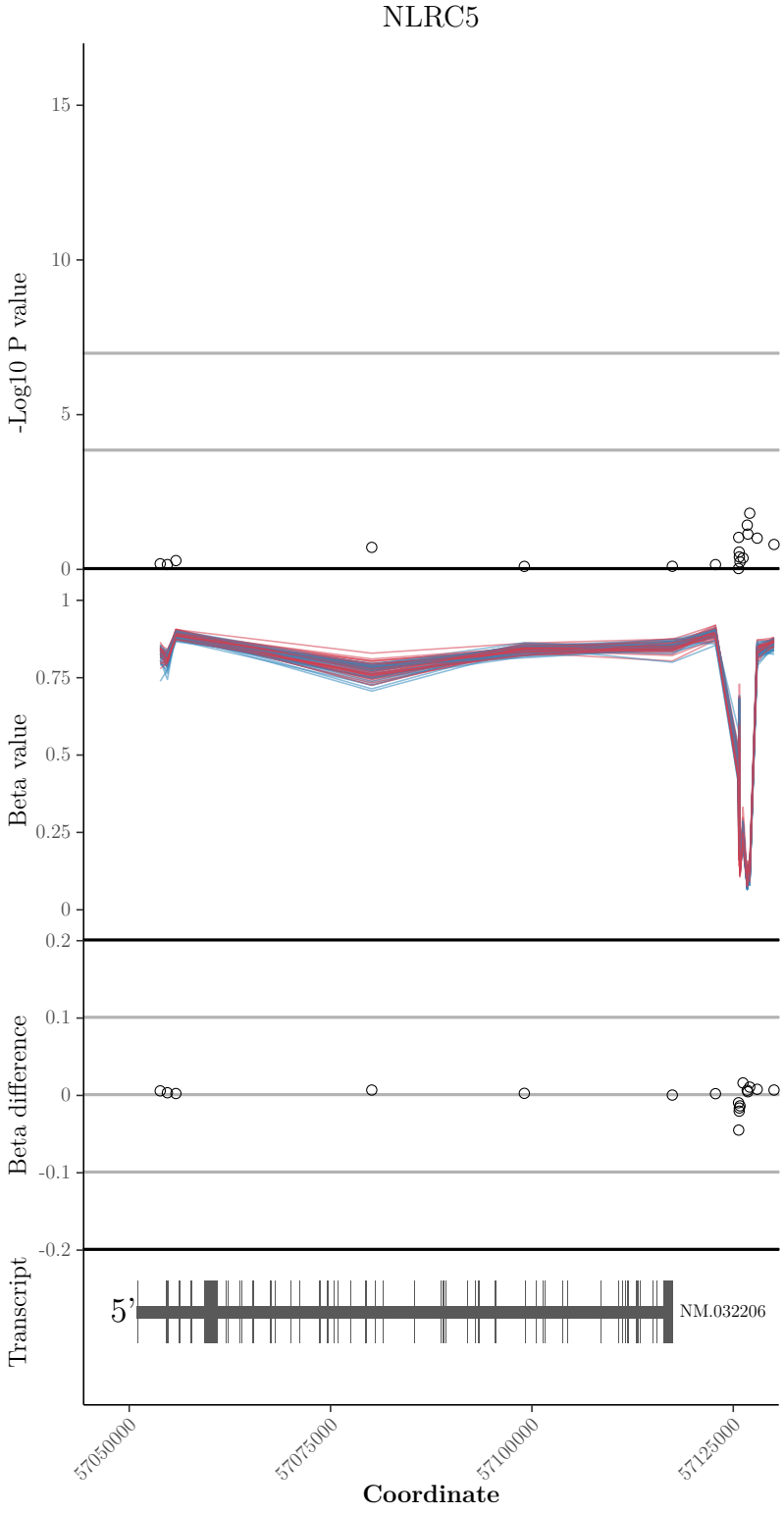


Figure 3.58: Methylation results for the NLRC5 region, including the DMR annotated to adjacent gene CETP (grey highlight)

ITGB2 ITGB2 (Integrin β2 or CD18) combines with ITGAL (Integrin αL or CD11a) to form LFA-1 (lymphocyte function-associated antigen-1), which is widely expressed on neutrophils, macrophages, T cells and B-cells; is important in leukocyte adhesion and may be altered in IBD [54, 242].

The autosomal recessive disorder type 1 leukocyte adhesion deficiency (LAD-I) is caused by a variety of mutations in the ITGB2 gene which reduce expression. The condition is characterised by recurrent infections of skin and mucosa, with reduced pus formation, and retarded wound healing. Severity is positively correlated with reduction in CD18 expression [242]. The level of expression of CD11a and CD18 is used as a predictor of severity, and in the most severe cases with less than 1% of neutrophils expressing these markers the average life span is 10 years compared to 40 years with 1–10% [172].

Although colitis in patients with LAD-1 may appear clinically and macroscopically similar to CD there is a major histological difference in that the underlying defect means there is no recruitment of neutrophils or formation of granulomas [645]. One study looked at 3 patients with inflammatory bowel disease with mild LAD-I and found that on top of the original mutation they had mutations in a subset of their T cells which produced a functional CD18 [646].

Efalizumab, a monoclonal antibody against CD11a, was tested in patients with moderate or severe CD in a small open-label pilot study and showed a positive effect on disease severity [292], however, the drug was withdrawn in 2009 due to cases of progressive multifocal leucoencephalopathy [292, 403].

Although these two effects act in opposition – loss of CD18 from a mutation causing colitis and treatment of colitis with an antibody against CD11a – this fits with the concept of a delicate homeostasis of immune activation at the intestinal epithelium. It should also be remembered that while the cases of IBD in patients with LAD-I may have appeared similar, they may well not share the same underlying abnormality which causes CD.

PRF1 Perforin 1 — or lymphocyte pore forming protein — is critical to the pro-apoptotic function of NK and cytotoxic T cells by forming pores across the membrane of target cells [501] allowing passage of granzymes such as GZMB which then induce apoptosis. Hypomethylation has been reported in CD4+ T cells from SLE patients [392], as opposed to the two hypermethylated DMRs in the promoter and gene body observed here and also in the 27k study by Nimmo *et al.* ($p=1.2\times10^{-7}$) [458]. The *PRF1* promoter is unmethylated in cytotoxic T cells which express it highly, but also in other tissues which do not express

PRF1 [679], and the expression of PRF1 is not straightforward in that NK cells have been shown to maintain a reservoir of *PRF1* and *GZMB* mRNA which is translated upon activation and is under microRNA regulation [325]. Mutations in *PRF1* are found in 20–50% of patients with the recessive condition familial haemophagocytic lymphohistiocytosis, and the similar complication of systemic onset juvenile idiopathic arthritis, macrophage activation syndrome [658]. Polymorphisms have been linked to a variety of other immune, autoimmune, and inflammatory conditions such as MS [95] and response to HIV infection [424].

TOLLIP Tollip has been shown to have a variety of immunomodulatory actions [94] and it has been proposed that failure to upregulate Tollip in response to inflammatory stimuli may be involved in the pathogenesis of inflammatory bowel disease [589].

SLC15A4 SLC15A4 is a transporter for a variety of substrates including peptides, sugars, and ions. As discussed in section 2.2.1, SLC15A4 is involved in transportation of MDP into the cell, and directly interacts with NOD2. It is additionally required for TLR7 and TLR9 signalling, and knockdown protects against DSS colitis in mice [319, 452, 491, 490].

ZBTB16 A zinc finger transcription factor also called promyelocytic leukaemia zinc finger (PLZF), due to the identification of a translocation causing a fusion product with retinoic acid receptor (RARα) in promyelocytic leukaemia. Implicated in several other cancers, adipogenesis [430], and T cell development [167]. Expressed in invariant natural killer T cells (iNKT), γδ T cells, NK cells, 5% of $CD8^+$ T cells, and 2% of $CD4^+$ T cells for a total of 10% of human PBLs [167].

DIABLO Also called SMAC (second mitochondria-derived activator of caspase), DIABLO is a pro-apoptotic mitochondrial protein which is released into the cytoplasm during apoptosis where it blocks inhibitor of apoptosis (IAP) proteins from performing their anti-apoptotic role of inhibiting caspases. DIABLO appears to be important in TRAIL-induced apoptosis [731]. TRAIL (TNF-related apoptosis inducing ligand, also called TNFSF10) exhibited Bonferroni significant CD-associated hypomethylation. The *DIABLO* DMR is hypermethylated and appears to be in the promoter region.

3.5 Conclusions

At the time of publication this work was the largest epigenome-wide association study in Crohn's disease, and has provided numerous targets which are forming the basis of ongoing work by our group. The results summarized below are of interest to IBD researchers and those planning epigenome-wide association studies for other complex traits.

1. Analysis strategies for EWAS
    (a) Combining the analysis of suitable cohorts is more powerful than separate analyses
    (b) Methylation results can be used to accurately characterise the constituents of blood in healthy and diseased individuals
    (c) Reference datasets may not be suitable for this purpose if the study population or tissue is unusual
    (d) Probe density must be controlled for when analysing genes or genomic regions
    (e) Familywise error rate correction methods are suitable for identifying a small group of 'hits' for further study, but are more conservative than the threshold derived from random permutations of this dataset and do not account for correlation between neighbouring probes
    (f) Equivalent results were obtained with any method of FWER correction
    (g) False discovery rate correction is suitable for identifying the larger numbers of results required to study broader patterns, though it may be insufficiently conservative as a basis for detailed or costly follow-up of individual results
    (h) β values and M values give very similar results, but other analytical techniques may be more appropriate than either
2. Small epigenome-wide association studies appear to be more powerful than predicted by simulations
    (a) The methylation findings are larger than those used to predict the group sizes needed to adequately power an EWAS
    (b) Groups of 17 would have 80% power to detect the strongest finding, and groups of 52 would have 80% power to detect all epigenome-wide significant findings

   (c) All 7 significant genes which were assayed by pyrosequencing replicated with group sizes of 20

3. The profile of epigenome-wide methylation changes in Crohn's disease

   (a) There is a pattern of hypomethylation and increased methylation variance in significant disease associated findings

   (b) Significant findings are predominantly found in regions of intermediate methylation, as opposed to most probes which are either highly methylated or mostly unmethylated

   (c) Differentially methylated regions contain probes with a distinct character compared to other probes of similar significance. They are substantially more likely to be hypermethylated, and associated with CpG islands

4. Highly significant disease-associated methylation differences

   (a) 65 CpGs were identified which survived Bonferroni correction for multiple testing

   (b) These include genes known to be involved in the pathogenesis of IBD and genes with biologically plausible links to IBD-related pathways

5. Methylation findings colocalise with GWAS risk loci

   (a) Methylation findings are significantly enriched in proximity to known genetic risk loci for IBD

   (b) Significant methylation probes are also significantly enriched in proximity to known genetic risk loci for several traits concerning cholesterol and lipid metabolism, as well as rheumatoid arthritis and type I and II diabetes

   (c) The analysis presented here offers a significantly improved approach to exploring this relationship

6. Inflammatory pathways and processes including ones known to be involved in IBD are enriched among genes with significant methylation changes

7. DNA binding factors are a possible cause of some disease-associated methylation changes

(a) Methylation probes are frequently designed to assay regulatory regions, and careful control is necessary

(b) There is evidence that a DNA binding motif for GLI1 and RUNX3 is enriched in significant probes, both of which have links to IBD

3.6 Supplemental

3.6.1 Pyrosequencing

Pyrosequencing is a method for rapidly and accurately sequencing short sequences of single stranded DNA. An initial PCR reaction incorporates biotin at the end of the target DNA strand. Biotin binds strongly to streptavidin in a manner that is resistant to a wide range of temperature, pH and denaturing conditions. The biotinylated DNA from each reaction is bound to streptavidin beads, which are then adhered to individual filters with a vacuum pump and dipped in sequential baths to denature and wash the products.

Once the vacuum is removed the beads fall into separate wells of a plate containing sequencing primer and buffer. This mixture is heated and allowed to cool, which anneals the sequencing primer to the biotinylated DNA. Once placed in the pyrosequencer, enzyme and substrate mixtures are added and then bases are added one at a time, causing a proportional release of light if they are incorporated. Any unincorporated bases are enzymatically degraded.

The sequencing reaction is detailed in figure 3.59 (page 171), and an example section of pyrogram is shown in figure 3.60 (page 172). This allows measurement of the relative proportions of various bases present at variable positions, such as those induced by bisulfite conversion of CpG sites. The measurement uses the surrounding sequence as a reference (and therefore the sequence must be known, and any variable bases taken into account), and is highly accurate and reproducible with a between measurement variation of <1%, but performs poorly with long target sequences (>200–300bp).

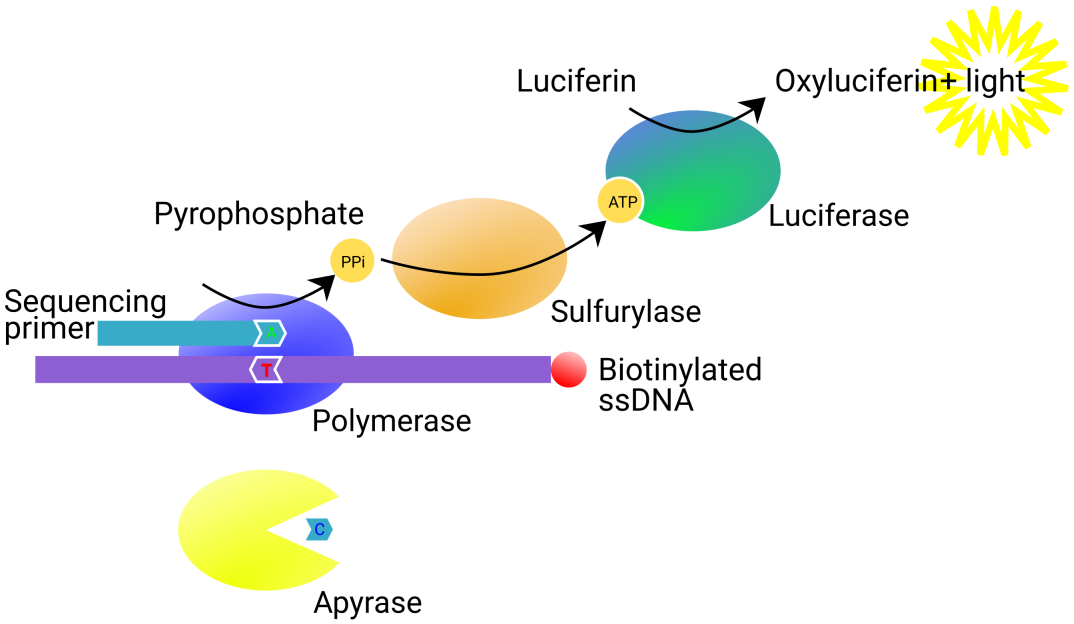


Figure 3.59: Pyrosequencing reaction: The sequencing primer is annealed to the target single-stranded biotinylated DNA molecule. dNTPs are released into the reaction mixture in a predefined order. Unincorporated dNTPs are degraded by apyrase, whereas dNTPs which are complementary to the target sequence are incorporated by DNA polymerase, resulting in the release of pyrophosphate (PPi), which is subsequently converted to ATP by ATP sulfurylase, which is used by luciferase to generate light in proportion to the amount of ATP present.

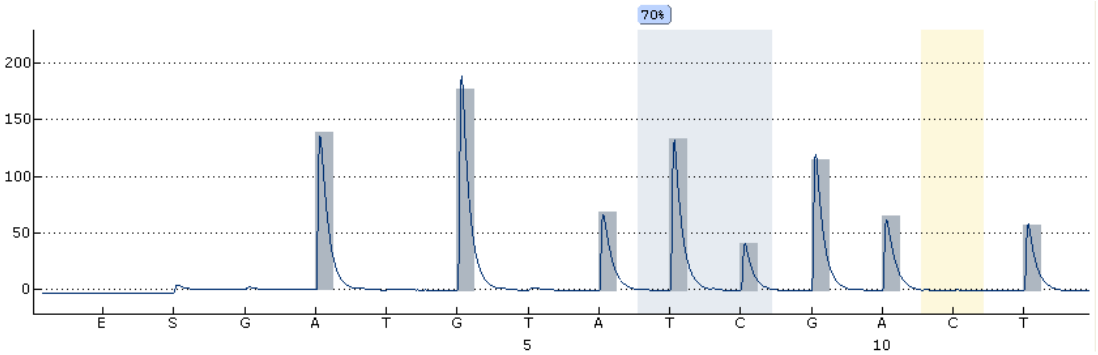


Figure 3.60: An example of a section of pyrogram from the pyromark Q24 software (Qiagen) for the DNA sequence `AAGGGATTYGGAT`. The background bars show the expected signal from the known sequence, and the line trace shows the level of recorded light signal. The blue shaded box containing `TC` is a methylation variable position, and the methylation (70%) is calculated by comparing the peak heights for `T` and `C` in reference to the fixed bases.
The dispensation order is shown along the x axis, positions with no corresponding bar for expected signal correspond to control bases to ensure specificity, to ensure full bisulfite conversion (highlighted yellow), and the initial addition of enzyme and substrate mixtures (E and S).
The colour of the box showing the measured methylation denotes the quality (blue - good, yellow - warning, red - failed). This is based on factors such as how well the measured fixed bases match the expected sequence, the shape and height of the peak traces, and whether there is any unconverted DNA detected at the bisulfite control base. Details of the quality assessment for each methylation reading are included in reports generated by the PyroMark Q24 software, with the colour coding allowing a rapid overview of sequencing results.

3.6.2 DNA motifs

Tables 3.24, 3.25, and 3.26 (pages 174–176) show DNA motifs in proximity to Bonferroni significant probes, FDR significant probes and DMRs respectively. The context of DNA motif `CCACGBVA` in Bonferroni significant probes is shown in table 3.23 below.

Gene	Chr	Sequence	Position	Context
ZBTB16	11	TTACGTGG	-2 – +5	intron 2/3–6
RTP5	2	TGGCGTGG	+ 694–701	intron 1/1
RUNX3	1	TGGCGTGG	- 710–717	§ intron 1–2/1–6
HEATR2	7	† TCACGTGG	- 550–557	§ intron 6/12
MPRIP	17	TGCCGTGG	- 734–741	§ intron 3/22–23
SLC15A4	12	TCACGTGG	- 40–47	intron 7/7
CNOT6L	4	† TCACGTGG	+ 242–249	§ intron 1/6–11
NMRAL1	16	TCACGTGG	- 3–10	intron 4/5
CBFA2T2	20	TTGCGTGG	+ 242–267	3′ UTR
PTDSS2	11	TGGCGTGG	- 172–179	intron 2/3–11
NLRC5	16	† TCACGTGG	+ 374–381	promoter
TNF	6	TGGCGTGG	-2 – +5	intron 3/3
TRPS1	8	TCACGTGG	-2 – +5	intron 4/5
NPDC1	9	TGGCGTGG	- 902–909	§ intron 1/8
-	15	TGCCGTGG	+ 151–158	no gene

Table 3.23: Sites near Illumina 450k probes with Bonferroni corrected p<0.05 which contain the DNA motif CCACGBVA. Sequence shown in sense orientation unless indicated with †, or no transcript is known. § possible alternative promoter for a shorter transcript. Context shows intron out of total number of introns, with a range if there are multiple transcripts. Position in bp relative to nominal probe target.

Consensus	Prevalence (%)	TF	Similarity (%)
CCACGBVA	**24.6**	**CREB3L1, CREB3**	**95.3**
		XBP1	**91.5**
		EGR4, EGR3	**89.8**
GBARRGGTCA	7.7	RARA	93.3
		RARA::RXRA	91.1
		NR2C2, NR2F1	88.7
CCCYTAG	**41.0**	**TFAP2B, TFAP2C, TFAP2A**	**90.2**
		PLAG1	**89.2**
		MZF1	**88.3**
		CTCF	**87.1**
HTGACACC	12.3	TBX2	90.1
		SMAD2::SMAD3::SMAD4	89.3
		TBX19, TBX20, TBX21	86.6
		MEIS3, MEIS2	86.1
		EOMES	85.7
		TBR1	85.4
		PKNOX1	84.5
GCGNCATYD	15.4	E2F2, E2F3	89.9
		YY2	88.6
AVCBTGTTT	**26.2**	**FOXA1**	**89.8**
CWAACBAAC	9.23	RREB1	89.7
CSGTTGA	18.46	MYBL2, MYBL1	89.5
		RXRA::VDR	88.3
		NR2F1	87.1
		RXRB, RXRG	86.0
		HR1H2::RXRA	85.6
		LEF1	84.6
RTTTGGCARC	10.8	NFIA	88.3
TSHAAVTTCA	16.9	RXRG, RXRB	88.0
		HNF4G	87.2
		NR1H2::RXRA	85.6
		NR2C2	84.1
CHCCTTTGT	9.2	KLF13	87.2
CTDTGNGGT	**32.8**	**GLI2**	**86.8**
		RUNX3	**86.5**
SDATCCANTC	12.3	NFYB, NFYA	86.2
DYTAGTGGT	12.3	CTCF	85.0
		NKX2-3	82.7
CTNCTGAGAA	12.3	STAT3	84.9
AAGATYGBNY	13.8	YY1	84.6
GCTCTTTDW	18.5	POU6F2	84.4
		MEF2B	83.4

Table 3.24: Shared DNA motifs near Bonferroni significant probes with the highest similarity to human transcription factors. Composite transcription factors are joined by two colons, family members with similar results are listed together. Motifs with prevalence greater than twice the mean of all discovered motifs are shown in bold.

Consensus	Prevalence (%)	TF	Similarity (%)
CCACGBVA	**35.6**	**CREB3L1, CREB3**	**95.3**
		XBP1	**91.5**
		EGR4, EGR3	**89.8**
CCCYTAG	**40.5**	**TFAP2B, TFAP2C, TFAP2A**	**90.2**
		PLAG1	**89.2**
		MZF1	**88.3**
		CTCF	**87.1**
AVCBTGTTT	**17.8**	**FOXA1**	**89.8**
CSGTTGA	7.17	NR2F1	89.0
		MYBL2, MYBL1	88.7
		RARA::VDR	86.7
		RXRB, RBXG	83.5
		NR1H2::RXRA	83.1
		LEF1	82.4
DTCTGACAMVT	1.44	SMAD2::SMAD3::SMAD4	88.3
TGGAACHT	5.94	HSF4, HSF1, HSF2	88.0
		TP63	85.8
CCTTNAGATG	1.97	RARA	88.0
CTDTGNGGT	**28.1**	**GLI2**	**86.8**
		RUNX3	**86.5**
GGAGRGGBCA	9.78	RARA::RXRA	86.3
		CTCF	83.4
		MZF1	82.3
		ZNF263	81.1
YATGCCTGGA	1.66	TP53	86.2
CDATTGYTVAM	1.22	MAFF, MAFG, MAFK	86.0
		NRL	85.6
		MAF::NFE2	83.7
CACKRTAAVC	3.23	PPARG	85.9
		ESR1	82.7
TTGCGGAWC	0.44	CEBPB, CEBPD, CEBPE, CEBPG	85.3
KAGGAGTTCC	1.31	BCLB6	84.4
GCTCTTTAT	1.40	MEF2B	84.4
		TBP	82.9
		MEF2D, MEF2A	82.7
		POU6F2	82.4
DTTAGTGGR	7.69	NKX3-2, NKX2-3	84.1
GCCCTKGNAAV	5.55	RFX5, RFX3, RFX2, RFX4	84.0
TATTWTCKTR	1.57	MEF2C	83.7
RTTTGYCAVC	4.63	NFIA	83.2
		SMAD2::SMAD3::SMAD4	80.8

Table 3.25: Shared DNA motifs near FDR significant probes with the highest similarity to human transcription factors. Composite transcription factors are joined by two colons, family members with similar results are listed together. Motifs with prevalence greater than twice the mean of all discovered motifs are shown in bold.

Consensus	Prevalence (%)	TF	Similarity (%)
TSHAAVRTCA	21.1	RXRB, RXRG, RXRA	95.9
		NR1H2::RXRA, RARA	95.6
		NR2C2	91.9
CCACGBVA	**47.4**	**CREB3L1, CREB3**	**95.3**
		XBP1	**91.5**
		EGR4, EGR3	**89.8**
CBCAGCVGGG	26.3	ZIC1, ZIC3, ZIC4	93.1
BDANCCANTC	15.8	NFYB, NFYA	90.8
ANCAWAATDT	26.3	BHLHE23, BHLHE22	90.3
		OLIG1, OLIG3, OLIG2	89.9
		NEUROG2	89.1
		TAL1::TCF3	86.8
CCCYTAG	**47.4**	**TFAP2B, TFAP2C, TFAP2A**	**90.2**
		PLAG1	**89.2**
		MZF1	**88.3**
		CTCF	**87.1**
CMGTTGA	21.1	NR2F1	88.9
		MYBL2	88.0
CGAACDC	31.6	RXRA::VDR	88.2
TGGCGGCWC	10.5	E2F3, E2F2, E2F7, E2F8, E2F1, E2F4	87.9
		ZBED1	87.2
GTGCNGHC	21.1	REST	87.7
CTDTGNGGT	31.6	GLI2	86.8
		RUNX3	86.5
AAGATSGYNCC	21.1	YY1	86.7
DWTAGNGGK	26.3	INSM1	86.6
		CTCF	86.2
CCCYRGNAAG	26.3	GLIS3, GLIS2:GLIS1	86.5
		BCL6B	84.0
GTAAAHAKAAC	15.8	FOXA1	85.5
YTCCTTTKT	26.3	SPIC	85.3
TGCCTGGAA	10.5	STAT3, STAT1	85.2
BTAAGYTG	26.3	NKX3-2	84.2
GGAGTCGBCR	21.1	EGR1	84.0

Table 3.26: Shared DNA motifs near DMRs with the highest similarity to human transcription factors. Composite transcription factors are joined by two colons, family members with similar results are listed together. Motifs with high prevalence are shown in bold.

Chapter 4

VMP1/MIR21

4.1 Introduction

The most significant individual probe from the analysis presented in chapter 3 was cg12054453 ($p=1.97\times10^{-15}$) located in the gene *VMP1* on chromosome 17. Five CpGs from *VMP1* were found to have significant differences in methylation following Bonferroni correction out of 65 total CpGs reaching this significance threshold — making up 8% of all Bonferroni significant probes. Additionally two neighbouring probes survived FDR, but not Bonferroni correction. Four consecutive significantly differentially methylated probes within *VMP1* met the criteria for a differentially methylated region (DMR, see sections 3.2.3 page 56 and A.3.5 page 262), and *VMP1* also lies within a GWAS risk locus for IBD [297] (see table 3.11, page 99).

Hypomethylation of this region in association with Crohn's disease is therefore the strongest novel finding of the work presented in chapter 3, and has prompted ongoing work to further characterise these findings and explore their implications. The following chapter will examine the context of the methylation findings within *VMP1*, propose possible biological changes that could cause or be caused by this hypomethylation, and demonstrate replication of the methylation findings in other IBD cohorts.

4.1.1 Context of VMP1 Illumina 450k results

Methylation results for all Illumina 450k paediatric results are shown in table 4.1 (page 179) and figure 4.1 (page 178). *VMP1* is relatively sparsely covered by the Illumina 450k chip compared to the average of 17 per gene region (promoter,

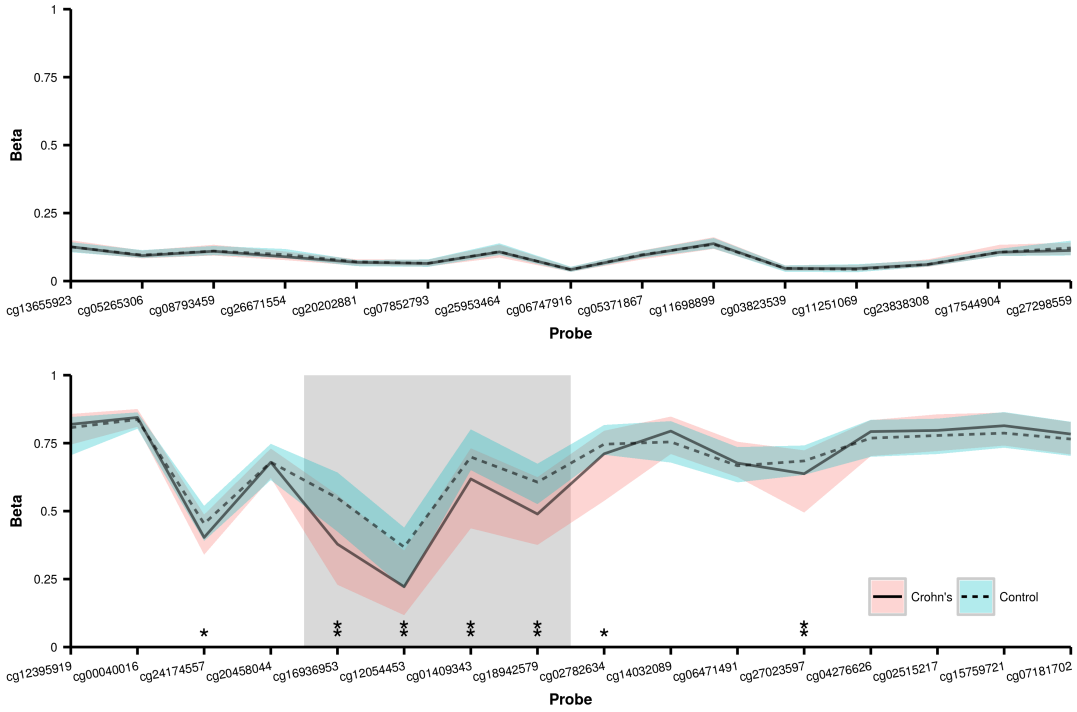


Figure 4.1: Methylation results for all 450k probes within the *VMP1* promoter (top) and gene body (bottom). The median value is represented by a line and the total range by the shaded ribbon. FDR (*) and Bonferroni (**) significance are indicated and the DMR is represented by the grey shaded area. X axis is not to scale. Adapted from Adams *et al.* [4, Figure 4C]

5'UTR, first exon, gene body, and 3'UTR) quoted by the manufacturer[3] and the high proportion of significant probes within the gene body is clearly visible in figure 4.1. This also shows that the promoter region of *VMP1* is largely unmethylated in both CD and control samples, with all of the significant results occurring within the gene body.

The significant methylation results within *VMP1* were defined as a differentially methylated region (DMR) by our criteria (page 56), which required consecutive FDR significant probes with a shared direction of methylation change within a variable distance threshold based upon normal CpG density for different genomic environments. The probes included in this DMR are indicated in figure 4.1. The position of all Illumina 450k probes within *VMP1* and their significance is shown in figure 4.2, where it can be seen that the significant methylation changes occur within the primary sequence of microRNA *MIR21*.

[3] `http://www.illumina.com/products/methylation_450_beadchip_kits.ilmn`

	Probe	Coord	P Value	β_{CD}-β_{HC}
	cg13655923	57784442	0.94	0.00
	cg05265306	57784459	0.16	0.00
	cg08793459	57784647	0.99	0.00
	cg26671554	57784674	0.01	-0.01
	cg20202881	57784779	0.61	0.00
	cg07852793	57784794	0.59	0.00
	cg25953464	57784812	0.85	0.00
	cg06747916	57784835	0.69	0.00
	cg05371867	57784854	0.47	0.00
	cg11698899	57785008	0.27	0.00
	cg03823539	57785039	0.79	0.00
	cg11251069	57785042	0.37	0.00
	cg23838308	57785056	0.56	0.00
	cg17544904	57785087	0.18	0.00
	cg27298559	57785523	0.20	-0.01
	cg12395919	57787151	0.54	0.01
	cg00040016	57833286	0.97	0.01
$*$	cg24174557	57903544	2.22×10^{-7}	-0.05
	cg20458044	57904327	0.02	0.00
$**$	cg16936953	57915665	2.65×10^{-13}	-0.17
$**$	cg12054453	57915717	1.97×10^{-15}	-0.15
$**$	cg01409343	57915740	1.58×10^{-8}	-0.08
$**$	cg18942579	57915773	8.23×10^{-10}	-0.12
$*$	cg02782634	57916643	5.63×10^{-6}	-0.04
	cg14032089	57917589	0.05	0.04
	cg06471491	57917974	0.65	0.01
$**$	cg27023597	57918262	5.42×10^{-8}	-0.05
	cg04276626	57918500	0.90	0.02
	cg02515217	57918600	0.78	0.02
	cg15759721	57918630	0.25	0.03
	cg07181702	57918682	0.54	0.02

Table 4.1: Paediatric Illumina 450k results for all probes across *VMP1* ordered by genomic coordinate, showing the p value and difference in mean β values between Crohn's and controls. ∗FDR significant, ∗∗Bonferroni significant

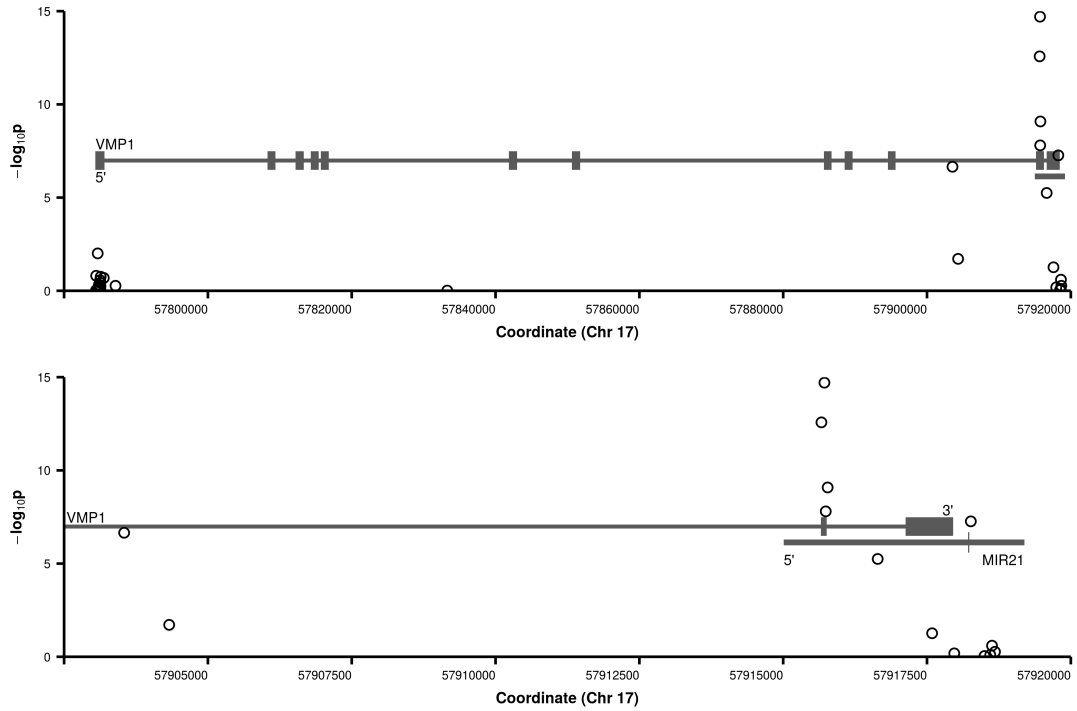


Figure 4.2: The position and significance of Illumina 450k probes with an overlaid schematic of *VMP1*, with the y position of introns demarking Bonferroni significance (top) and a zoomed view showing the significant probes and the position of *MIR21* (bottom). Adapted from Adams *et al.* [4, Figure 4A & B]

4.1.2 VMP1

Localisation was first mapped in rats using EGFP-tagged VMP1, showing it as a transmembrane protein located in the Golgi apparatus, endoplasmic reticulum (ER) and in vacuoles, whose formation it promotes [163]. VMP1 has also been shown to be expressed in the ER in *Drosophila*, where it is known as TANGO5 [36], *C. elegans* [621], and *Arabidopsis* [161], in mammalian autophagosomes [527] and in the plasma membrane of cancer cells [549]. Homologues are absent in fungi but present in some protists implying a loss during early fungal evolution [89]. Its high conservation means human VMP1 is able to replicate the function of its homologue in *C. elegans* (EPG-3) [621], and rat VMP1 in *Dictyostelium discoideum* [89].

Expression and induction VMP1 is expressed in rat acinar cells during acute [163, 647] and chronic pancreatitis [295] where it has been shown to promote the formation of intracellular vacuoles and cause apoptotic cell death. It is expressed in foetal rat pancreas but is then undetectable by northern blot until induction of pancreatitis [163, 647].

In WBN/Kob rats, which spontaneously develop chronic pancreatitis, RT-PCR detected *VMP1* mRNA from age 8 to 20 weeks, peaking at 12 weeks, coinciding with maximal vacuole formation and the onset of pancreatitis. However, levels were undetectable by 24 weeks, at which point vacuolisation and apoptosis were still evident, although reduced [295].

In control animals, *VMP1* transcripts of 1.9 and 2.7kb [163, 295] were highly expressed in intestine, kidney, ovary and placenta, moderately expressed in liver, lung, stomach, thymus, brain, and testis and, slightly expressed in thyroid and retina. A band of 3.5kb was also observed in intestine, liver, lung, kidney, stomach, thymus, ovary and placenta [163].

A 30 minute period of ischaemia to rat kidney caused a large increase in expression of *VMP1* in that tissue, but *VMP1* remained undetectable in pancreas in those animals, and increased expression following induction of acute pancreatitis was restricted to pancreas [163].

Autophagy VMP1 has also been shown to be induced by autophagy triggers such as starvation and rapamycin [305]. Autophagy is highly relevant to Crohn's disease (see section 2.2.2, page 14) with many of the most strongly CD-associated genes such as *NOD2* and *ATG16L1* being involved in autophagy (see table 2.2,

page 15). VMP1 localises to the site of autophagosome formation in the ER independently of other autophagy proteins including ULK1, WIPI-1, DFCP1 and ATG16L [290], but is critical for the proper formation and maturation of autophagosomes in *Dictyostelium discoideum* [89, 90, 91], mouse embryonic fibroblasts [290], *Chlamydomonas reinhardtii* [611], *C. elegans* [621], and humans.

VMP1 interacts with other members of the Class III PI3K complex (Beclin-1[4], Vps34, Vps15, and ATG14L), via direct interaction between the Beclin-1 BH3 domain and the 20aa C-terminal hydrophilic domain of VMP1 where it is required for recruitment of the PI3K complex to the autophagosomal membrane, which is in turn responsible for recruitment of other autophagy proteins such as ATG16L1 and LC3 [435, 305, 91, 290]. Additionally, VMP1 interrupts the interaction of BCL2 with the Beclin-1 BH3 domain, which removes the autophagy-diminishing influence of BCL2 [435], and directly interacts with other proteins such as PT53INP2 [464].

VMP1 is required for elongation and closure of the isolation membrane [331, Figure 8], reduced expression of VMP1 leads to disrupted autophagosomal maturation, with persistent omegasomes (an early stage in autophagosome formation) and delayed onset of autophagy [91, 621], whereas overexpression causes autophagy and apoptosis regardless of starvation or rapamycin [163, 527, 648].

Other functions Apart from the vacuole formation observed in pancreatitis [647], VMP1 is required for formation of the contractile vacuole in *Dictyostelium discoideum*, an organelle responsible for osmoregulation [89]. Other proposed roles for VMP1 include cell adhesion [549], cell death [90, 647], ER integrity, membrane traffic, organelle organisation, [89], protein secretion [36], cytokinesis, cell cycle, morphology in *Chlamydomonas reinhardtii* [611], and zymophagy — a protective response to pancreatitis leading to autophagy of zymogen granules, involving USP9x and p62 (SQSTM1) [221]. In cancer decreasing VMP1 expression correlates with increased malignancy, metastasis, invasion, and proliferation and a reduction in median survival [232, 549].

[4] ATG6 in yeast

4.1.3 MIR21

Chapter 2 contains background information on microRNAs:

Section	Page	Contents
2.4.3	28	general background
2.5.1	34	findings in IBD
2.5.2	36	as biomarkers
2.5.3	41	as therapeutic targets

As shown in figure 4.2 (page 180), the primary transcript of mir21 (*Pri-mir21*) overlaps with the 3′ end of *VMP1*. This region is highly conserved, exhibits DNase I hypersensitivity and is associated with the promoter-associated histone marks H3K4Me1 and H3K4Me3 [220]. There is evidence of genomic instability in this region with tandem duplication in breast cancer leading to VMP1-RPS6KB1 fusion and worsened prognosis in these patients. This fusion transcript was detectable at very low levels in healthy controls, possibly indicating trans-splicing [285]. The region is also a site of HPV integration in cervical cancer (fragile site 17B) [556] and a frequent site of duplications in neuroblastomas [544].

A promoter region for *MIR21* has been described around the penultimate exon of *VMP1* and contains a TATA box and binding sites for transcription factors including STAT3, p53, C/EBP, NFI, and AP-1 [199, 552]. Alternative independent transcription start sites have been found in the same region, with some evidence of cell type specificity [87, 391, 476, 199, 516], or even 135kb upstream — overlapping the upstream region, 5′ UTR, first exon and first intron of *VMP1* [411]. The primary transcript of mir21 extends beyond the 3′ UTR of *VMP1* with the mature sequence also located beyond the 3′ UTR. There is clear evidence that mir21 has promoters which are independent from *VMP1* [199]; however, an alternative, independently regulated source of mir21 has been described from VMP1 transcripts which are extended beyond the normal polyadenylation signal and processed to produce mature mir21 — '*VMP1-mir21*' [517].

Multiple *VMP1* transcripts (1.9, 2.7 and 3.5kb, section 4.1.2) were found in rats [163, 295] although alternative splicing of *VMP1* has not been demonstrated and these may represent the products of *pri-mir-21* or *VMP1-mir21* cleavage by Drosha [517]. The varying levels of the different rat *VMP1* transcripts between tissues, and *pri-mir21*:*VMP1-mir21* ratios between cell lines may indicate this alternative source of mir21 is regulated in a tissue specific manner [517, 295, 163].

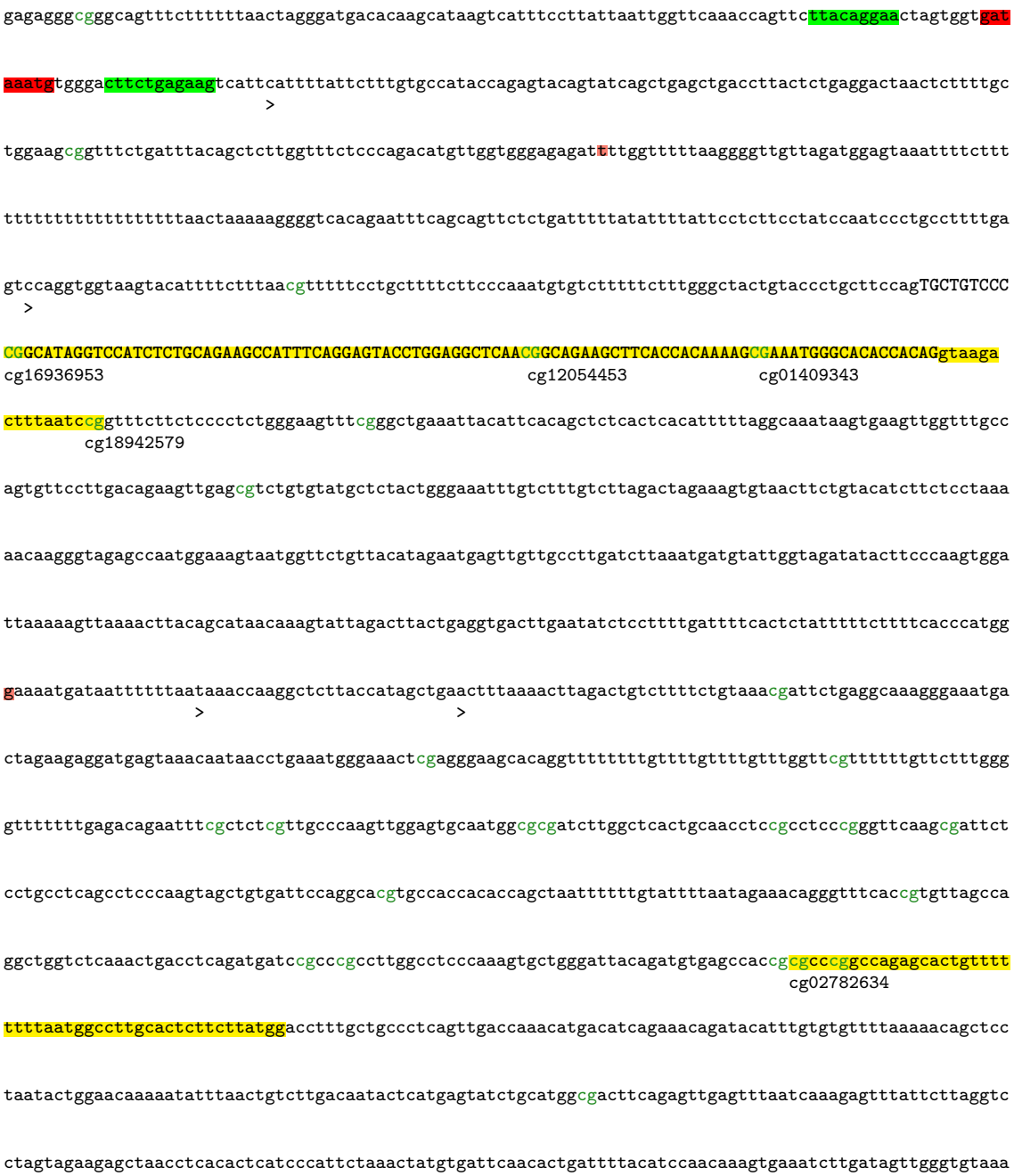


Figure 4.3: Position of VMP1 probes, CpGs, SNPs with MAF>0.01 in the European population, TATA box, STAT3 binding sites, reported mir21 TSS (>) [199, 476, 87], `introns`, and **`EXON 11`**

The production of a mature microRNA from a transcript which overlaps with a gene transcript does decrease the expression level of that gene as the remaining gene transcript has a naked 3′ end and is not processed normally [87, 517]. However, allowing the processing of a luciferase mRNA containing a microRNA only decreased the luciferase twofold [87], and Ribas *et al.* showed that knocking down Drosha, preventing the processing of *VMP1-mir21*, led to increased *VMP1* mrna levels [517].

If an ongoing inflammatory condition such as CD drove both mir21 and *VMP1-mir21* and the ratio of the two varied in some cell types due to the presence of certain genetic or epigenetic variants affecting processing, then this could impact autophagy via levels of VMP1.

Function Mir21 was one of the earliest described microRNAs and has been implicated in numerous cancers, including lung, breast, ovarian, pancreatic, prostate colorectal and IBD-associated colorectal cancer [303, 304, 396, 199, 347, 697, 516, 585], by reducing apoptosis and affecting growth, transformation, invasion, and metastasis, and therefore prognosis [471, 457, 697, 98]. Mir21 has also been implicated in numerous non-neoplastic conditions including myaesthenia gravis [505], SLE [492], type 1 diabetes [492], cardiac conditions (ACS, MI, aortic stenosis, cardiac arrest), MS, HCV infection, as well as both Crohn's disease and ulcerative colitis [234]. Mir21 is also involved in T cell differentiation and development [394, 104, 552, 396]. Mir21 is overexpressed in IBD [565, 704, 600, 703, 725], and in mouse colitis models. However, *MIR21* knockout in mice has been shown to be protective against DSS-induced colitis [565], but deleterious in TNBS-induced colitis and T cell transfer colitis [701].

Regulation VMP1 and mir21 do share some regulators, such as androgens, AP-1 [517], and possibly IL-4 [534, 535]. Other known regulators include PMA/AP1 [199], IL-6 (via STAT3 [391]) STAT3, BCL6 [552], NFκB [736], androgen receptor [516] and TLR4 [471]. Expression of mir21 in healthy rectal mucosa has also been shown to increase in response to a high red meat diet [277].

Mir21 targets Mir21 has 2522 targets with a "good" mirSVR score of <-0.1 — from an algorithm based on experimental data correlating the characteristics of mir binding sites with changes in expression after microRNA transfection. The algorithm includes factors such as predicted accessibility, flanking sequences, seed

type, position within the target 3′ UTR, and conservation [58][5].

Though intragenic microRNAs do sometimes regulate expression of their host genes, *VMP1* does not contain a mir21 binding site, however a feedback mechanism could involve intermediaries such as STAT3 which has binding sites in the *VMP1* promoter (figure 4.3) and has 4 mir21 binding sites in its 3′ UTR. NFIB is both a negative regulator of mir21 expression, and directly negatively regulated by mir21 [199].

There was no significant enrichment of mir21 targets among Bonferroni significant methylation results (15.7% vs 13.2% overall, p=0.69), but there are many genes relevant to IBD among those mir21 has been shown to target, or with strong but untested mir21 binding sites. Figure 4.2 shows some of the best characterised mir21 targets and targets of greatest relevance to IBD.

Having identified the *VMP1* exon 11 / pri-mir21 region as the most significant result from the genome-wide methylation analysis, work was undertaken to measure methylation of this same region in independent IBD cohorts and in other diseases, and to examine expression.

[5]Data available at www.microrna.org

Target	Comments
PDCD4	A highly studied mir21 target, PDCD4 (programmed cell death 4, or neoplastic transformation inhibitor protein) is a tumour suppressor which has been proposed to explain a substantial portion of mir21 oncogenic and prognosis worsening behaviour in cancer.
SPRY2	Homologue of *Drosophila* gene *Sprouty* with roles in epidermal growth factor and MAP kinase signalling. Also proposed as a link between mir21 and cancer [680].
PTEN	Involved in ROS-induced NOD2 and IL-6-STAT3-mir214 signalling, both leading to NFκB expression [625, 499, 471].
SMAD7	A negative regulator of TGFβ involved in T_H17 differentiation [471]. A phase 2 trial of oral anti-SMAD7 oligonucleotide has shown positive results in CD [437]
NFIB	A negative regulator of mir21 expression, which is itself negatively regulated by mir21 [199]
VIM	Vimentin is involved with membrane localisation of NOD2, which is disrupted by CD-associated *NOD2* variants, reducing AIEC clearance [590].
STAT3	STAT3 is known to upregulate mir21 expression in response to IL-6 and IL-21, and competes with negative regulator BCL6 for a shared binding site in the *MIR21* promoter [552].
CARD8	Interacts with NOD2 and has CD-associated polymorphisms (see section 2.2.1) [672, 421]
IL10	Anti-inflammatory cytokine, discussed in relation to SOCS3 and SBNO2 in section 3.4.5
TLE1	A transcriptional corepressor which opposes NFκB, directly interacts with NOD2 and has CD-associated genetic variants [459].
LRRK2	Autophagy gene which functions with NOD2 in response to bacteria, and has genetic links to CD, ankylosing spondylitis and leprosy (section 2.2.2).

Table 4.2: Continued overleaf

Target	Comments
PTPN2	Autophagy gene involved in response to TNFα, IFNγ, and intracellular bacteria, with CD-associated genetic variants (section 2.2.2).
Methylation results	Bonferroni significant probes INPP4B, POLK, MPRIP, NLRC5, PRKCE, and FKBP5
Epigenetic	DNMT3B, an IBD GWAS [297] risk locus, responsible for new DNA methylation (with DNMT3A, also a GWAS risk locus [194]), and DNMT1 indirectly via RASGRP1 [478]. Also targets HDAC8, HDAC9, and SIRT5.
Interleukins	As well as IL10 (above) IL1B, IL1RAP, IL9, IL12A, and IL22
TNFRSF	Multiple members of the TNF receptor superfamily including TNFRSF6B, TNFRSF10B, TNFRSF10D, TNFRSF11B, TNFRSF17, with varied roles in immunity, apoptosis and NFאB signalling, often as decoy receptors. E.g. TNFRSF10B, 10D and 11B are decoy receptors for TNFSF10 (TRAIL), which had Bonferroni significant methylation changes ($p=3.28\times10^{-10}$).
TGFβ	Effects on differentiation and proliferation [329] through targeting multiple TGFβ related products such as TGFB1, and TGFB2, and TGFβ receptors TGFBR2, and TGFBR3.

Table 4.2: Transcripts with experimentally validated or predicted mir21 binding sites, focusing on products with proven or suspected importance to CD or IBD

4.2 Materials & methods

4.2.1 Pyrosequencing

Sample collection Initial pyrosequencing replication in 20 adults with Crohn's disease and 20 age and sex-matched controls was performed as described in sections 3.2.2 (page 54) and 3.3.8 (page 117), further details on pyrosequencing are shown in section 3.6.1 (page 170).

Samples were obtained for genetic research (LREC 2000/4/192) and inclusion in the IBD-BIOM and IBD Character studies from patients recruited from gastroenterology clinic at the Western General Hospital and healthy volunteers. The three large adult pyrosequencing replication cohorts were obtained from this same source, with the initial cohort of 20 CD and 20 controls being included in the first large replication cohort. Summary demographics are shown in tables 4.3 and 4.4.

Samples of whole blood DNA from groups of 20 patients with ankylosing spondylitis, rheumatoid arthritis, and systemic lupus erythematosus were kindly provided by the Rheumatological research group[6]. Demographics are shown in table 4.6.

	Initial		Extended	
	CD	HC	CD	HC
Age range	20.8–57.1	20.7–57.0	18.7–75.7	18.5–74.7
Median	34.2	34.2	37.7	37.3
Male(%)	45	50	44	45
NOD2 WT (%)	70	85	78	84
Non-smoker (%)	100	100	100	100
White European (%)	100	95	98	90

Table 4.3: Demographics for the initial 20v20 pyrosequencing replication (figure 3.41, page 118) and the first large pyrosequencing replication (figure 4.4, page 196). *NOD2* WT denotes wild type *NOD2* status, defined as not carrying any of the R702W, G908R, or fs1007insC CD-associated variants.

[6]Institute of Genetics and Molecular Medicine, University of Edinburgh

	Initial			Extended	
Montreal %	-L4	+L4		-L4	+L4
L1	15	0	L1	15	1
L2	55	0	L2	41	1
L3	15	5	L3	30	3
L4	5		L4	2	
Oral	5		Oral	1	
Unknown			Unknown	15	

Table 4.4: Montreal classification (see table 2.5, page 45 for explanation) of disease location in the the initial adult 20v20 replication cohort (figure 3.41, page 118) and the first large adult replication cohort (figure 4.4, page 196).

	IBD	CD	UC	Con	SC	HC
Age						
Range	18.0–79.1	18.0–79.1	18.1–74.5	18.3–75.0	18.3–69.5	18.5–75.0
IQR	26.4–48.2	26.2–47.3	26.8–48.7	27.5–43.0	25.7–36.7	30.8–49.6
Median	33.2	32.8	34.0	32.8	29.2	38.2
Female (%)	37.0	37.4	36.7	53.7	70.5	36.7
Smoking (%)						
Current	20.4	30.8	10.0	19.0	24.6	13.3
Ex	30.9	24.2	37.8	28.9	23.0	35.0
Never	47.0	42.9	51.1	51.2	52.5	50.0

Table 4.5: Demographics for the second adult replication cohort. Con: all controls, SC: symptomatic controls (IBS), HC: healthy controls (volunteers).

		Age		
	N	Range	Median	Male(%)
AS	18	22–90	62.5	62.5
RA	19	18–82	67.0	29.2
SLE	19	28–86	53.0	8.3

Table 4.6: Rheumatology cohort demographics

Bisulfite conversion The initial adult cohort (n=40) for replication by pyrosequencing was bisulfite converted with the column-based EZ DNA methylation kit (Zymo Research Corporation, Irvine, CA, USA. Cat no. D5001), the other larger replication cohorts were bisulfite converted using 96-well EZ-96 DNA Methylation kits (*ibid.* Cat no. D5003) with 500ng input DNA per sample.

Pyrosequencing The first large adult replication cohort was analysed using a PyroMark Q96 ID system (Qiagen, Hilden, Germany), all other cohorts were analysed using a PyroMark Q24 system. The assays (table 3.4, page 55) were designed using PyroMark Assay Design Software (version 2.0.1.15, Qiagen) and ordered from Sigma-Aldrich (St. Louis). Methylation values and sequence quality were exported from the manufacturers software and analysed in R, with significance determined by Wilcoxon rank-sum test.

Samples were run in duplicate, results with any quality warnings were discarded and repeated, analysis was performed on the mean of two methylation readings within 2% absolute methylation. Where the variance between repeats was greater than 2% a further set of duplicate reactions was assayed, if these two readings and one of the original readings were all within 2% (probably due to a pipetting error or contamination of one of the original samples) the mean of these three was used, otherwise the sample was excluded.

Further explanation of pyrosequencing and the factors involved in sequence quality assessment can be found in section 3.6.1 (page 170).

Assay		Primer
1	F:	TTTTGTGGGTGTGGTAAGAG
	R:	[Btn]-ACCAACCTACCTACAATACC
	S1:	GGTAAGAGGAAGAATGA
	S2:	GGAATTTGTTATTTAGGAAT
	S3:	GTTGTTTAGAAGAAGGGAA
2	F:	[Btn]-ATTAGTTTTTTAAAGTGTTGGGATTATAGG
	R:	ACAAACCTCATACCCTCAACTA
	S1:	CTAAACACTTATAAATATCCCC
	S2:	AAAAAAAATTAATTTCTAAAAATA
3	F:	AGTGTTTAGTTGAGGGTATGAG
	R:	[Btn]-AAAAACTCAAAATTTCTAATTCCCAACA
	S:	AGTTGAGGGTATGAG
DMR	F:	TGGGTTATTGTATTTTGTTTTTAGTGTTG
	R:	[Btn]-ACTAACAAACCAACTTCACTTATTTAC
	S:	GTATTTTGTTTTTAGTGTTGTT

Table 4.7: Pyrosequencing primers for three assays covering the VMP1 5′ UTR CGI (top) and the previously described DMR assay (bottom, see also table 3.4, page 55). Forward (F), reverse (R), and sequencing (S) primers. [Btn]: 5′ biotinylation, 0.05µmol synthesis scale and HPLC purification

4.2.2 qPCR

MIR21 primary transcript expression was analysed in patients with CD (n=44) recruited from gastroenterology clinics and endoscopy lists (Western General Hospital, Edinburgh, UK), and compared with healthy volunteers (n=22). Written informed consent was obtained prior to sample collection[7].

Blood was collected in a 9ml K3 EDTA Vacuette tube (Greiner) and RNA was extracted from 1.5ml of this within 2 hours at 4°C using a QIAamp RNA Blood Mini Kit (Qiagen). RNA was stored at -80°C until cDNA conversion with SuperScript Vilo cDNA Synthesis Kit (Invitrogen). Primers were ordered from Sigma-Aldrich (table 4.8) and qPCR was performed with a Rotor-Gene 6000 (Corbett life science) and DyNAmo Flash SYBR Green (ThermoFisher). Ct values were measure in duplicate reactions, and analysed by the δδCt method.

Normalisation was performed as described by Vandesompele *et al.* [657], with sequential exclusion of reference genes. Final normalisation was performed against TBP.

Gene		Primer
pri-mir21	F:	ATGGGCTGTCTGACATTTTGGTA
	R:	CATTGGATATGGATGGTCAGATGA
TBP	F:	TGCCCGAAACGCGAATATA
	R:	TTTCTTGCTGCCAGTCTGGA
SDHA	F:	AGGGCATCTGCTAAAGTTTCAGA
	R:	GATTCCTCCCTGTGCTGCAA
ACTB	F:	GCCAGCTCACCATGGATGAT
	R:	AATCCTTCTGACCCATGCCC
GAPDH	F:	CGACCACTTTGTCAACTCA
	R:	AGGGAGATTCAGTGTGGTG

Table 4.8: pri-mir21 and control qPCR primers

4.2.3 Microarray expression data from Noble *et al.*

Reanalysis of raw microarray data previously generated by our group, and published by Noble *et al.* [461, 462], was used to examine *VMP1* and *pri-mir21* expression in inflamed and non-inflamed biopsies from controls (56 vs. 17), Crohn's disease (48 vs. 51), and ulcerative colitis (65 vs. 64). see Section 3.2.1, page 52 for further details.

[7]LREC 06/S1101/16, LREC 2000/4/192

4.3 Results

4.3.1 Replication of methylation findings

Replication by pyrosequencing As shown in figure 3.40 (page 116) there was sufficient DNA to replicate one finding by pyrosequencing using the paediatric DNA. *VMP1* probe cg12054453 was chosen and replicated well with an r^2 of 0.95.

Secondly as shown in figure 3.41 (page 118), initial replication was performed in a cohort of 20 adults with CD and 20 healthy controls, on seven significant results including *VMP1* which replicated with p=0.0038. As presented in Adams *et al.* [4], replication was also successfully performed in a cohort of 172 adults with a p value of 6.6×10^{-5} (shown in figure 4.4, page 196).

In the same cohort methylation of a CpG island in the presumed *VMP1* promoter region was also assessed by pyrosequencing. Longer PCR products were produced due to low availability of primer binding sites and higher densities of CpGs in the CpG island compared to the regions assayed to confirm significant Illumina 450k findings. This allowed the use of multiple sequencing primers with the same PCR product to assay a wider area. As shown in table 4.7 (page 192) there were in effect six assays. These were used on 95 of the 172 adult samples above, but there was no difference between disease and controls and all samples measured <2% methylated (results not shown), further pyrosequencing of this area was therefore not undertaken.

Two further replications by pyrosequencing have since been performed using the same assay of the mir21 region, both successfully demonstrating the same hypomethylation in association with Crohn's disease, and also for IBD overall ($p=1\times10^{-8}$, n=141[8] & $p=1\times10^{-6}$, n=266[9]) [663].

Replication by Illumina 450k Our group have performed two further Illumina 450k studies. The first by NTV as part of the IBD-BIOM project also found hypomethylation of the same probes within *VMP1* among the strongest results (CD $p>1.15\times10^{-13}$, IBD $p>2.15\times10^{-11}$, both Holm corrected) [663].

Pre-publication analysis of the Illumina 450k data from the IBD Character study shows significant hypomethylation at the same five probes as the paediatric data for IBD in all samples ($p=2.40\times10^{-7}$–3.66×10^{-20}) and in samples from centres other than Edinburgh ($p=3.56\times10^{-6}$–7.73×10^{-15}), in CD ($p=8.70\times10^{-9}$–

[8]Assay design: ATA, Patient selection: ATA, Experimental work: NTV & ATA
[9]Assay design: ATA, Patient selection: NTV, Experimental work: WTCRF

1.58×10^{-17}) and in UC (p=1.49×10^{-7}–4.01×10^{-15}). *VMP1* was the strongest result for UC, third strongest for CD (after two *SBNO2* probes), and was the strongest result for IBD in the non-Edinburgh samples. Further results from the IBD Character project are shown in appendix A.1, page 253.

Methylation in rheumatological conditions *VMP1* methylation was assayed in groups of 20 samples from patients with ankylosing spondylitis (AS), rheumatoid arthritis (RA), and systemic lupus erythematosus (SLE), using the same assay as in figure 4.4. The results are shown in figure 4.5 (page 197). None of the other tested conditions were significant differently methylated compared to healthy controls. Crohn's disease samples were hypomethylated compared to rheumatoid arthritis (p=0.016) and SLE (p=0.028), but not ankylosing spondylitis (p=0.084).

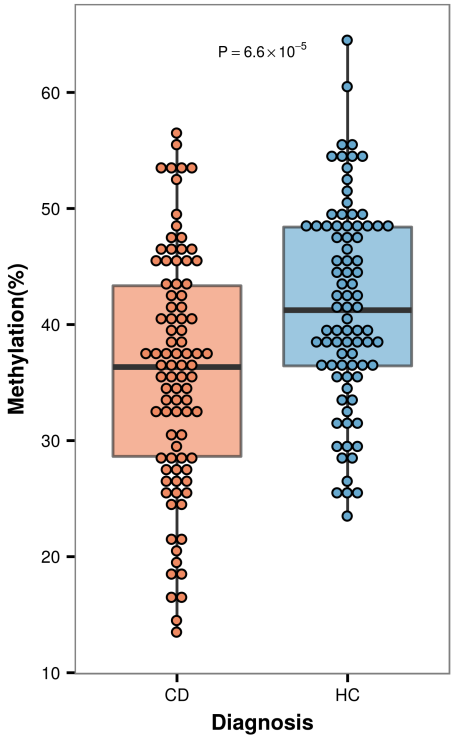


Figure 4.4: Box and whisker plot showing pyrosequencing of *VMP1* corresponding to probe cg16936953 in 87 adults with CD and 85 healthy adult controls, as reported in Adams *et al.* [4, Figure 5A]. Whiskers extend up to 1.5 × IQR from the 1st and 3rd quartiles, with all results also plotted as points.

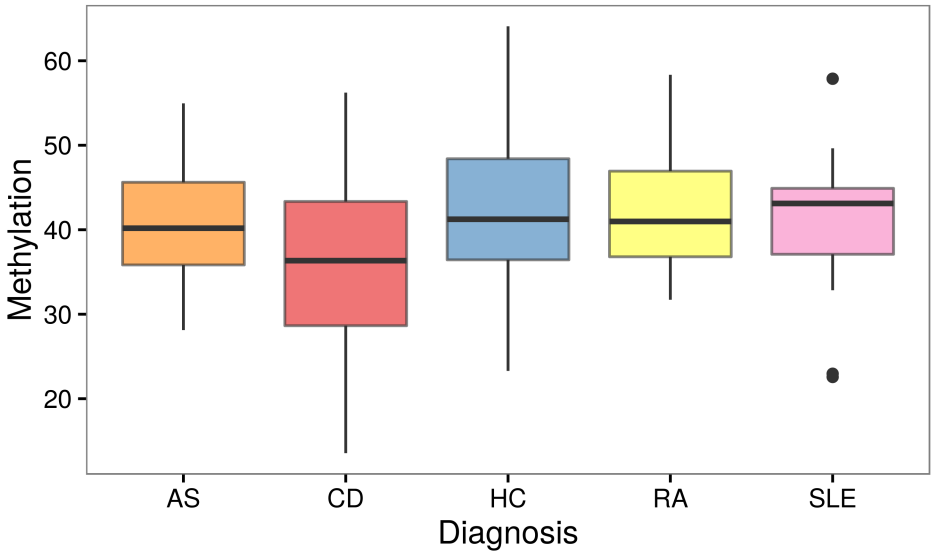


Figure 4.5: Box and whisker plot showing pyrosequencing of *VMP1* corresponding to probe cg16936953 in ankylosing spondylitis, rheumatoid arthritis, and systemic lupus erythematosus, compared with Crohn's disease and controls. Whiskers extend up to 1.5 $\times$ IQR from the 1st and 3rd quartiles, with any remaining results plotted as points.

4.3.2 Expression

Analysis of blood and biopsy microarray data Data generated by our group and published in Noble *et al.* [462, 461] was used to assess *VMP1* and *pri-mir21* expression in inflamed and non-inflamed biopsies from patients with CD and UC, and from healthy controls and controls with non-IBD inflammation. These results are shown in figure 4.6.

VMP1 showed increased expression in CD samples with inflammation compared to CD samples without inflammation ($p=2.6\times10^{-3}$), which was not observed in either UC patients or controls.

Both CD and UC samples showed a significant increase in *pri-mir-21* expression ($p=1.4\times10^{-6}$ and 5.1×10^{-7} respectively) in response to inflammation. There was a trend toward increased *pri-mir-21* expression in inflamed controls but this was non-significant, possibly as a result of small number of inflamed controls.

The only differences in baseline expression were between CD and UC, with *MIR21* having a slightly lower expression ($p=0.003$) and *VMP1* having slightly higher expression ($p=0.03$) in CD.

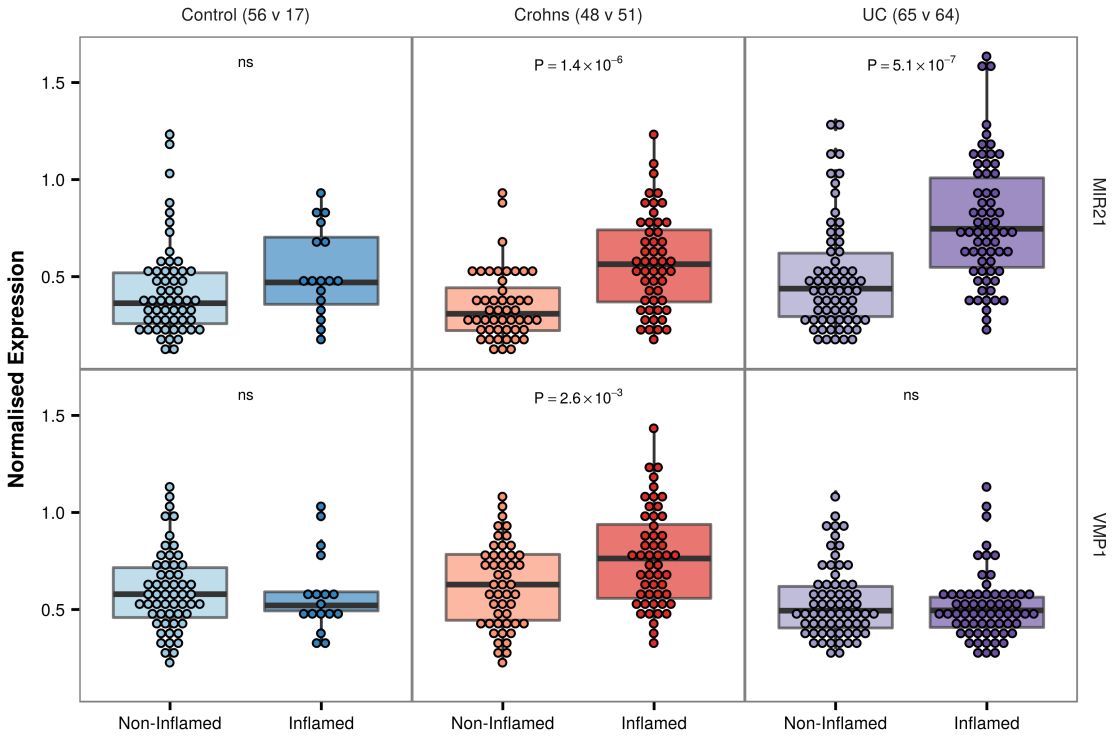


Figure 4.6: Box and whisker plots showing microarray data for VMP1 and MIR21 expression in non-inflamed and inflamed biopsies from controls, Crohn's disease, and ulcerative colitis as reported in Adams *et al.* [4, Figure 5C]. Whiskers extend up to 1.5 × IQR from the 1st and 3rd quartiles, with all individual results also plotted as points. P values for inflamed vs non-inflamed by Wilcoxon rank sum test by diagnosis and target shown in each facet. Sample numbers given at the top. Reanalysis of data from two papers by Noble *et al.* [461, 462], see section 3.2.1, page 52 for further details.

Primary transcript expression *Pri-mir-21* expression was assessed in whole blood by qPCR (figure 4.7). There was a 41.5% increased expression of the primary transcript (p=0.008) in Crohn's disease (n=44) compared to healthy controls (n=22).

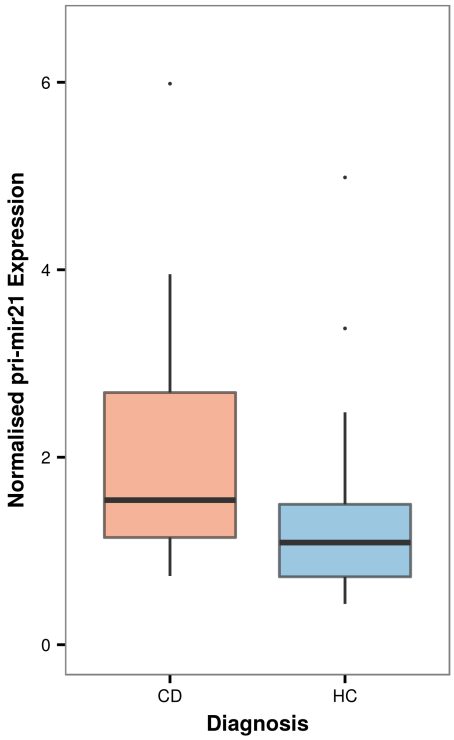


Figure 4.7: Box and whisker plot showing pri-mir21 expression in whole blood measured by qPCR, as reported in Adams *et al.* [4, Figure5B]. Whiskers extend up to 1.5 × IQR from the 1st and 3rd quartiles, with any remaining results plotted as points. P=0.008 (Wilcoxon rank sum test)

4.4 Discussion

Methylation Hypomethylation of the region surrounding the 11th exon of *VMP1* has now been demonstrated in 3 cohorts using the Illumina 450k platform and 3 cohorts using pyrosequencing. A subgroup analysis of the IBD Character study replicated the findings even when excluding all Edinburgh samples, leaving a primarily Scandinavian cohort.

The finding has not been reported in other IBD EWAS studies (table 2.4, page 33), although as discussed in section 2.5 the variety of platforms, analytical techniques, and tissues studied makes comparisons between studies difficult.

Expression We have demonstrated increased expression of *pri-mir-21* in blood by qPCR and inflamed biopsy specimens by microarray. These findings are consistent with a growing number of reports of increased mir21 expression in IBD [704, 600, 703, 725, 565]. Unfortunately, before the paediatric methylation results and their initial replication drew our attention to mir21 and other microRNAs, the RNA extraction kits we used for biobanking did not adequately preserve small RNAs, preventing immediate assessment of mature mir21 levels.

We have since added microRNA collection to our biobanking protocols and work by RK is ongoing examining microRNA expression in leucocytes and serum in IBD. Initial unpublished observations confirm that mir21 is highly expressed in IBD. On the basis of the preliminary results presented here our group obtained funding for further investigations[10] into *VMP1* and *mir21*. This ongoing work, particularly focused on *VMP1*, *mir21*, and their relationships with autophagy and IBD, covers many of the natural follow-up experiments which would have fitted well into this thesis, but my role in the work is insufficient for their inclusion here.

The microarray expression data presented in Adams *et al.* [4] and in figure 4.6 (page 199) assumes that the mRNA corresponding to *VMP1* intronic sequence within the *pri-mir21* sequence measures *pri-mir21*, and that the probe corresponding to *VMP1* sequence measures *VMP1*. However, the existence of an independently regulated alternate *VMP1* transcript extending beyond the usual polyadenylation signals (*VMP1-mir21*) as shown by Ribas *et al.* [517] which is processed to produce mature mir21 complicates this picture. Further work should be undertaken to assess if the *VMP1-mir21* transcript is found in the blood or bowel in IBD, and whether DNA methylation in the 11th exon of *VMP1* impacts

[10] work by KRO and ERN

its regulation.

Genetics The *VMP1* DMR is within a genetic risk locus for IBD (table 3.11, page 99) [297], within which no causal variants have been identified. A shortcoming of this experiment was that funds were not available for sequencing the surrounding region in detail or to perform genotyping arrays. This has been addressed in following studies. Work currently underway within our group is sequencing the entire *pri-mir-21* sequence and flanking regions in an adult cohort of >700, in whom *VMP1* methylation will also be assayed[11]. Pre-publication results from NTV [12], with further confirmation in the IBD Character project (appendix A.1, page 253) show a strong correlation between *VMP1* DMR methylation and local SNPs rs8078424 and rs10853015, which are in LD with the marker SNP (rs1292053) for the GWAS risk locus described by Jostins *et al.* [297].

Biological relevance of MIR21 to IBD Overexpression of mir21 has now been repeatedly shown in IBD [704, 600, 703, 725, 565] and in mouse colitis models [565, 701], and in response to diet [277].

Mir21 has been implicated in many diseases (section 4.1.3) and is clearly not specific to IBD, but part of fundamental inflammatory pathways. This does not detract from its relevance to IBD, which is clear when considering the broad mechanism of action of many of the most potent anti-inflammatory drugs in IBD and their widespread utility in other conditions.

Analysis of *VMP1* exon 11 methylation in rheumatological samples shows significant differences between Crohn's disease and the other complex inflammatory conditions rheumatoid arthritis and systemic lupus erythematosus, which themselves were indistinguishable from healthy controls. Ankylosing spondylitis could be concluded to be intermediate between Crohn's disease on one hand and RA, SLE and controls on the other. Given the established overlap between AS and CD in genetic and other risk factors, and their co-occurrence within families and individuals (5–10%) [669, 661, 619, 518], this finding does appear interesting; however, there are too many uncontrolled variables to draw any conclusions from these data. A specifically designed study of adequate size would be required, capable of matching or correcting for gender, age, disease activity, medications, smoking status and ethnicity, where rheumatological, IBD and control samples were all collected and processed identically.

[11] KRO & ATA
[12] Ventham *et al.* [663]

A large number of genes contain conserved mir21 targets in their 3′ UTRs, including many of high relevance to IBD (table 4.2, page 188). However, only a small proportion have been experimentally confirmed, and the complexities of feedback and regulatory systems is barely explored. This could explain the varying effects of mir21 knock-down in different mouse colitis models [565, 701]. There is a clear impetus to undertake further work examining the downstream effects of altered mir21 expression in IBD, and the potential therapeutic avenues in disrupting this.

Although the DMR within *VMP1* appears to be related to mir21, relationships between intragenic microRNAs and their host genes can be complex and interrelated, and there is some evidence for this with mir21 and *VMP1* (see sections 4.1.2 and 4.1.3). The essential role of VMP1 in autophagy, a core pathway in CD pathogenesis, strengthens the impetus to include *VMP1* in the ongoing work by our group rather than focusing solely on mir21. Work by KRO is currently ongoing investigating the role of VMP1 and mir21 in autophagy in Crohn's disease.

Conclusions In conclusion, this work presents novel evidence of a highly significant and replicable hypomethylated region around the 11$^{\text{th}}$ exon of *VMP1* associated with both Crohn's disease and ulcerative colitis. Additionally these data support the previously reported finding of increased expression of mir21 in IBD. Future work to examine the downstream effects of altered mir21 expression in IBD, and the potential therapeutic implications of disrupting this would be highly valuable. Further work by NV, KRO and AA is examining this region, known to exist within a large GWAS risk locus for IBD, for genetic variants associated with IBD and with the methylation findings described here.

Chapter 5

RPS6KA2

5.1 Introduction

The second most significant individual probe from the combined paediatric Illumina 450k analysis presented in chapter 3 was cg17501210 ($p=4.47\times10^{-15}$) within the gene *RPS6KA2*. This chapter presents replication of this methylation finding in three adult cohorts comprising a total of 264 patients with Crohn's disease, 174 patients with ulcerative colitis, and 268 healthy controls, and replication of the finding in adult IBD patients and healthy controls using Illumina 450k data from the IBD-BIOM study [663][1] and the IBD-Character study (pre-publication, see appendix A). Secondly, to look for functional differences associated with the highly replicable methylation change associated with IBD, expression data for *RPS6KA2* in biopsies, whole blood and separated blood cells are presented, and cellular starvation is used to look for an effect of *RPS6KA2* knockdown on autophagy.

Background RPS6KA2 (ribosomal protein S6 kinase A2, also known as RSK3) is a serine/threonine kinase in the highly related 90kDa ribosomal S6 kinase family (RSK1–4). Additional family members and related enzymes are shown in table 5.1; note that RPS6KB1 neighbours *VMP1* within the same GWAS risk locus (see section 4.1.3).

The RSK family is included in the larger AGC kinase family, which have highly conserved catalytic kinase domains closely related to those found in protein kinase A, G, and C (PKA, PKG, and PKC) [488].

Members of the RSK family are differentiated from other AGC kinases in

[1] Work by NTV

that they contain two, highly-conserved, kinase domains which were separately acquired rather than duplicated [296]. The N-terminal kinase domain (NTKD) matches other AGC kinases and phosphorylates substrates; the C-terminal kinase domain (CTKD) matches the calcium/calmodulin dependent protein kinase family and activates the NTKD [488, 139]. These kinase domains are separated by a linking region of approximately 100aa with essential regulatory domains, with greater variability found in the N- and C-termini which may relate to specificity [593, 526]. The overall sequence similarity for *RSK1–4* is 73–80% [593, 526], upto 90% in the kinase domains [359], with RSK3 being most closely related to RSK (84% homology), and containing a unique N-terminal sequence not found in other RSKs [734].

Two longer isoforms of RSK3 have been described, though expression has been reported as <1% of the shorter, first described isoform [63]. There is little information about the function of these alternate transcripts, however the unique N-terminal sequence of RSK3, possibly required for nuclear localisation [734], would be disrupted in the longer transcripts.

Activation and function The RSK family were identified when searching for the enzymes responsible for phosphorylating ribosomal protein S6. Although RSKs were found to perform this function *in vitro*, and derive their name from it, they do not play a significant role in RPS6 phosphorylation *in vivo* where it is primarily performed by the 70kDa S6 kinases (S6K1 and S6K2) [526].

RSK1–3 are present in the nucleus and cytoplasm, with nuclear translocation following activation [109, 139] by ERK1/2[2] together with PDK1, another AGC kinase. The related proteins MSK1 and MSK2 are activated by p38[3] in addition to ERK1/2 and PDK1. Phosphorylation of certain residues in the RSK protein as well as autophosphorylation are also required for enzymatic activity [488, 593, 19, 139]. RSK1–3 are also activated by PMA[4], other MAPK activated kinases (MK2 and MK3 [727]), and general MAPK signalling — hence the alternative name MAP kinase-activated protein kinases [139]. RSK4 is different from the other RSK enzymes in that it is expressed at low levels (5–10% of other isoforms), does not require PDK1 for activation, and basal levels of ERK are sufficient for maximal activation [160].

The RSK family are ubiquitously expressed, but with tissue specific variation

[2]MAPK3 and MAPK1
[3]MAPK14
[4]phorbol 12-myristate 13-acetate

in isoform proportions [63, 359, 526]. *RSK1* mRNA is most highly expressed in lung, bone marrow and T cells; *RSK2* in T cells, lymph nodes and prostate; and *RSK3* in lung, brain, spinal cord, and retina [359].

Mutations in *RSK2* leading to lost kinase activity cause Coffin-Lowry syndrome, which features severe mental retardation, facial and digital abnormalities, kyphoscoliosis, and cardiac defects [634] with similar findings in RSK2-null mice [497]. These findings are not obviously linked to the tissue distribution of *RSK2* mRNA, and indeed, apart from RSK4 which is present at similar levels in all tissues, constitutively active and possibly regulated at the level of transcription, the RSK family appear to be regulated by activation rather than expression [160, 359].

Although protein motifs recognized by RSKs have been characterised [139, 367], as shown with RPS6, predicted or *in vitro* kinase activity may not represent significant *in vivo* activity. Additionally, isoform specific information is sparse due potential redundancy between family members [727], and mainly concerns RSK1 and RSK2 [359].

However RSKs have been shown to be involved in the regulation of widespread processes and pathways including MAPK signalling, transcription, apoptosis, glycogen metabolism, cytoskeletal structure, and cell growth, proliferation, motility and survival [526, 727, 734, 583], through phosphorylation of numerous targets [19, 526, 359]. These include IκB Kinase, leading to NFκB activation [479, 323], SRF and c-FOS, JUN, CREB, and NFAT3 (Reviewed by Lara *et al.* [359]).

Genetic variants within the RSK family have been linked to increased risk of colon cancer (*RPS6KA1*, *RPS6KA2*, and *RPS6KB2*), and variants within *RPS6KA2* linked to increased risk of rectal cancer [580], possibly mediated by diet [579].

RPS6KA2 and smoking An association has been reported between *RPS6KA2* methylation at probe cg22717080 and smoking (uncorrected p=1.68×10^{-7}, 1.7% hypomethylation in current smokers, 0.3% hypomethylation in ex-smokers)[5] [636]. This CpG is in the same intron as the CpG found to be hypomethylated in CD (10.7kb distal) but no significant association between CD and methylation at this probe was found in the paediatric Illumina 450k data, and there was also no correlation between the methylation at cg22717080 and cg17501210, the probe were we found CD-associated methylation changes. The reported smoking-associated methylation change displayed no correlated change in expression, and was not

[5]Bonferroni corrected p=0.06 for m=355628

Gene	Enzyme	Class	Location
RPS6KA1	RSK1	90kDa	1p36.11
RPS6KA2	RSK3	90kDa	6q27
RPS6KA3	RSK2	90kDa	Xp22.12
RPS6KA4	RSKB/MSK2	90kDa	11q13.1
RPS6KA5	RSKL/MSK1	90kDa	14q32.11
RPS6KA6	RSK4	90kDa	Xq21.1
RPS6KB1	S6K1	70kDa	17q23.1*
RPS6KB2	S6K2	70kDa	11q13.2

Table 5.1: RSKs and related enzymes. *Within the same region as *VMP1/mir21*.

found to be correlated with changes in blood cell counts [636].

Context of 450k results The Illumina 450k results for the RPS6KA2 region are shown in figure 5.2 on page 210. There was a single significant probe (cg17501210), which is located in the first intron of the short transcript (2^{nd}–3^{rd} of longer transcripts, see figure 5.1). This was the second most significant probe overall, with a p value 4.47×10^{-15} and an 0.11 decrease in β values from controls to Crohn's.

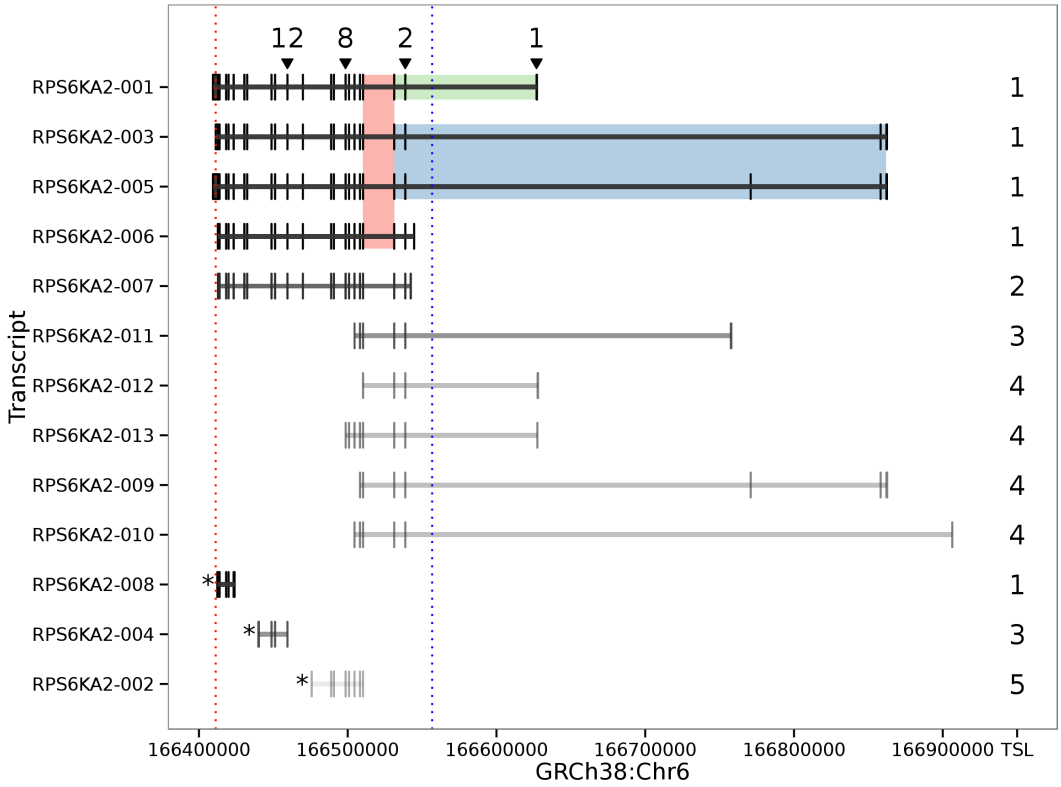


Figure 5.1: All experimentally validated and predicted isoforms of RPS6KA2, named as per ensembl.org. TSL: transcript support level (ensembl.org and UCSC genome browser), *: non-coding, dotted lines: microarray probes for expression (red) and methylation (blue), solid colour blocks show qPCR targets (for TSL 1 transcripts only) against all transcripts (red), transcript 1 (green) and long transcripts (blue). Numbered arrows show exons of antibody binding sites.

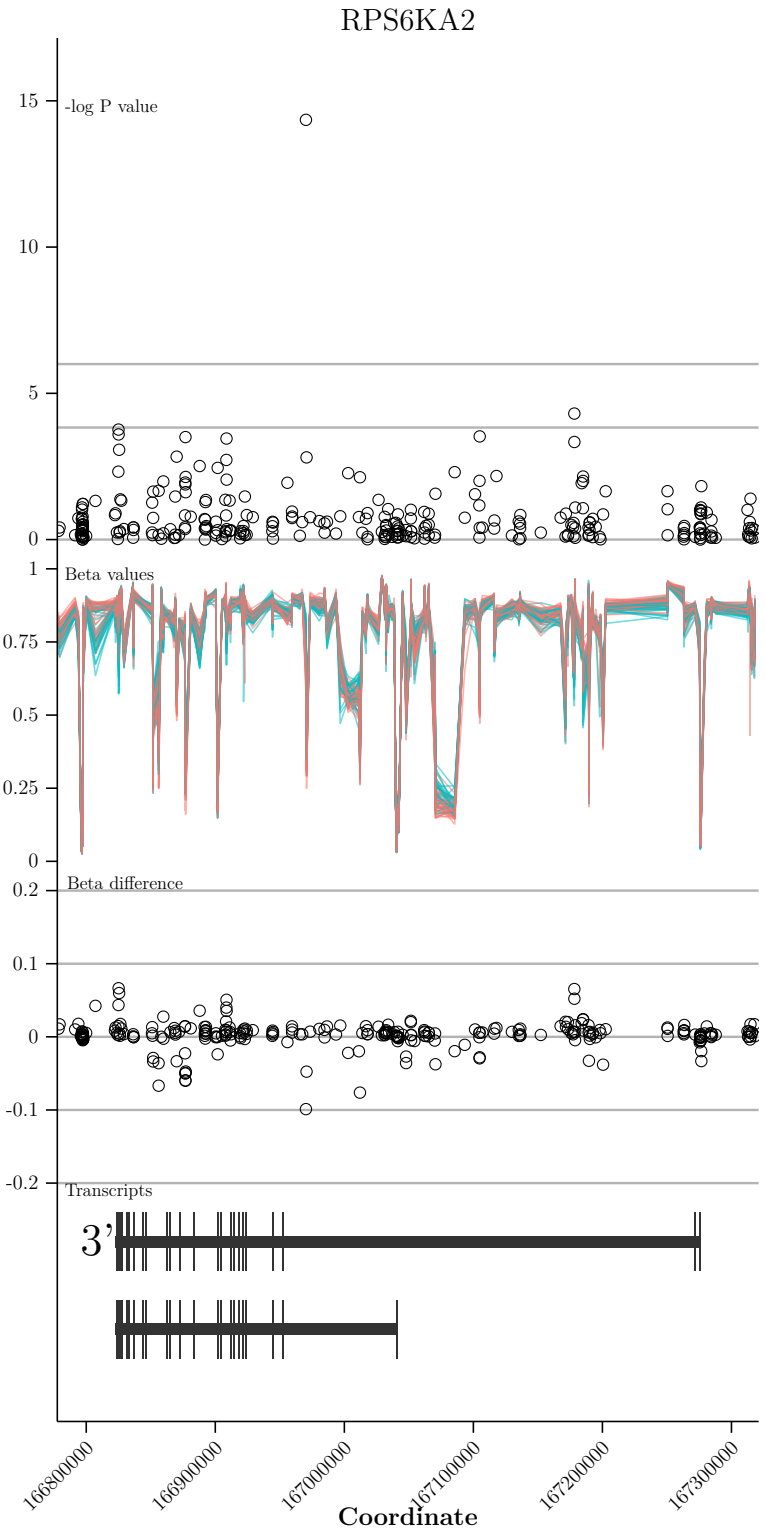


Figure 5.2: Context of RPS6KA2 Illumina 450k results. From top: 1) -$\log_{10}$ p values, with lines indicating Bonferroni corrected $p<0.05$ and FDR<0.05. 2) Beta values for each sample, red - CD, blue - control. 3) Mean beta value for CD relative to mean control. 4) Gene transcripts.

5.2 Materials & methods

5.2.1 Sample collection

Initial pyrosequencing replication with 20 adults with Crohn's disease and 20 age and sex-matched controls was performed as described in section 3.3.8 (page 117).

Cohorts 1–3 (figs. 5.3 & 5.4, pages 216 – 217, and table 5.4, page 215) samples were obtained for genetic research (LREC 2000/4/192) and inclusion in the IBD-BIOM and IBD Character studies from patients recruited from gastroenterology clinic at the Western General Hospital and healthy volunteers. Summary demographics are shown in sections 3.2 and 4.2

Bisulfite conversion The initial adult cohort for replication by pyrosequencing was bisulfite converted using the column-based EZ DNA methylation kit (D5001), the samples for larger replications were performed using 96-well EZ-96 DNA Methylation Kits (Zymo Research Corporation, Irvine, CA, USA. Cat no. D5003) with 500ng input DNA per sample.

Pyrosequencing was performed using the PyroMark Q24 System (Qiagen, Hilden, Germany). The assay (table 5.2) was designed using PyroMark Assay Design Software (version 2.0.1.15, Qiagen) and ordered from Sigma-Aldrich (St. Louis). The assay covers 3 CpGs, two of which are measured by the Illumina 450k Methylation probe cg17501210, and one adjacent CpG (figure 5.11, page 227). Methylation values and sequence quality were exported from the manufacturers software and analysed in R, with significance determined by Wilcoxon rank-sum test.

Samples were run in duplicate, results with any quality warnings were discarded and repeated, analysis was performed on the mean of two methylation readings within 2% absolute methylation. Where the variance between repeats was greater than 2% a further set of duplicate reactions was assayed, if these two readings and one of the original readings were all within 2% (probably due to a pipetting error or contamination of one of the original samples) the mean of these three was used, otherwise the sample was excluded.

Further explanation of pyrosequencing and the factors involved in sequence quality assessment can be found in section 3.6.1 (page 170).

Primer	Sequence
Forward:	GGTGGAGTTTATTGGAAGGTTGTG
Reverse:	[Btn]-ACAAAATCCCTCTAAATCCAACTATCT
Sequencing:	TGGGTGGTTTATTTAGAAT

Table 5.2: Pyrosequencing primers for RPS6KA2 [Btn]: 5′ biotinylation, 0.05µmol synthesis scale and HPLC purification

Assay		Sequence
All	F:	GCCACCCTAAAAGTTCGGGA
	R:	GGGGTGATTCACTTCTGCCA
Short	F:	AAAAGATCGAGGTGGAGCCT
	R:	CCAGGAACACCTTTCCATAGGAT
Long	F:	AGGAAGTCGCGCTCCAAGAG
	R:	TCCTCACCAGGAACACCTTTC

Table 5.3: qPCR primers for RPS6KA2, colours correspond to figure 5.1 (page 209)

5.2.2 qPCR

Patients with CD were recruited from gastroenterology clinics and endoscopy lists (Western General Hospital, Edinburgh, UK), and compared with healthy volunteers. Written informed consent was obtained prior to sample collection[6].

Blood was collected in a 9ml K3 EDTA Vacuette tube (Greiner) with PBMCs and granulocytes isolated within 1 hour of venepuncture. PBMCs were obtained by density centrifugation with Ficoll-Paque PLUS (GE Healthcare) as per manufacturers instructions. After the PBMC layer had been removed, the remaining Ficoll was discarded and the pellet was lysed with red cell lysis buffer to extract granulocytes. RNA from whole blood was obtained using 'QIAamp RNA Blood Mini' kits during the Ficoll centrifugation, RNA from granulocytes and PBMCs was extracted immediately after separation using 'miRNeasy Mini' kits (Qiagen).

RNA was stored at -80°C until cDNA conversion with SuperScript Vilo cDNA Synthesis Kit (Invitrogen). Primers were ordered from Sigma-Aldrich (table 4.8) and qPCR was performed with a Rotor-Gene 6000 (Corbett life science) and DyNAmo Flash SYBR Green (ThermoFisher). Primers are shown in table 5.3, and control primers in table 4.8, 193. Normalisation was performed as described by Vandesompele *et al.* [657], with sequential exclusion of reference genes. Final normalisation was performed against TBP.

5.2.3 Western blot

HEK293 cells expressing LC3-GFP were plated at 2×10^5 cells per well in 6 well plates and after 24 hours treated with RPS6KA2 siRNA (Dharmacon J-004663-06) or control siRNA as per manufacturers instructions for 60 hours. Cells were then either returned to DMEM + 10% FCS (fed) or treated with HBSS for 3

[6]LREC 06/S1101/16, LREC 2000/4/192

or 6 hours (starvation). One well for each combination of RPS6KA2 siRNA and starvation condition was treated with 200nm bafilomycin, and another was untreated.

Cells were then scraped, protein extracted and RPS6KA2 and LC3 assayed by western blot, and slides for each condition prepared for microscopy.

Probing for RPS6KA2 used polyclonal rabbit IgG from R&D systems (AF893) against phosphorylated serine 218 at 1:1000 dilution.

Confirmation of reduced RPS6KA2 expression was confirmed by western blot and qPCR. Cells were plated and after 24 hours incubated with RPS6KA2 siRNA as above without subsequent starvation. Cells were harvested at 0, 36, 60, and 72 hours incubation, with four replicates per time point. RPS6KA2 protein and mRNA expression was assayed as above.

5.2.4 Microarray expression data from Noble *et al.*

Reanalysis of raw microarray data previously generated by our group, and published by Noble *et al.* [461, 462], was used to examine expression of *RPS6kA2* and other RSK family members in inflamed and non-inflamed biopsies from controls (56 vs. 17), Crohn's disease (48 vs. 51), and ulcerative colitis (65 vs. 64). see Section 3.2.1, page 52 for further details.

5.3 Results

5.3.1 Pyrosequencing

As shown in figure 3.41 (page 118), initial replication was performed in a cohort of 20 adults with CD and 20 healthy controls, on seven significant results including *RPS6KA2* which replicated with p=4.4×10^{-5}.

As with *VMP1* (section 4.3.1, table 4.3), pyrosequencing of the same assay was performed in the 20 vs 20 cohort extended to n=172 ("cohort 1"), a second adult cohort ("cohort 2")[7] largely overlapping with the Illumina 450k study undertaken by NTV [663], and a third adult cohort ("cohort 3")[8]. The 2nd and 3rd cohorts contained both Crohn's disease and ulcerative colitis patients, and both healthy volunteers and symptomatic (IBS) controls.

Significant methylation differences were found between Crohn's disease and controls across the three CpGs measured in all three cohorts (see figure 5.3, page 216). In cohorts 2 and 3 there were no discernible differences between healthy controls and IBS despite appreciable differences in age, gender and smoking, with IBS being younger, more female, and more likely to smoke (see table 4.5, page 190). These two groups were therefore combined into a single control group in each cohort. Results were not significantly different between cohorts, and so a combined analysis is also shown in figure 5.4 (page 217).

CpG2 is the nominal target of the Illumina 450k probe, CpG3 is also contained within the probe, and CpG1 is just adjacent. Figure 5.11 (page 227) shows the location of the CpGs in relation to the probe and nearby SNPs and CpGs.

	1	2	3	Total
CD	84	83	97	264
HC	76	116	76	268
UC	-	77	97	174
Total	160	276	270	606

	Coordinate
CpG 1	166970255
CpG 2	166970252
CpG 3	166970248

Table 5.4: Cohort sizes and CpG coordinates for *RPS6KA2* replications by pyrosequencing

[7]Assay design: ATA, Patient selection: ATA, Experimental work: ATA
[8]Assay design: ATA, Patient selection: NTV, Experimental work: WTCRF

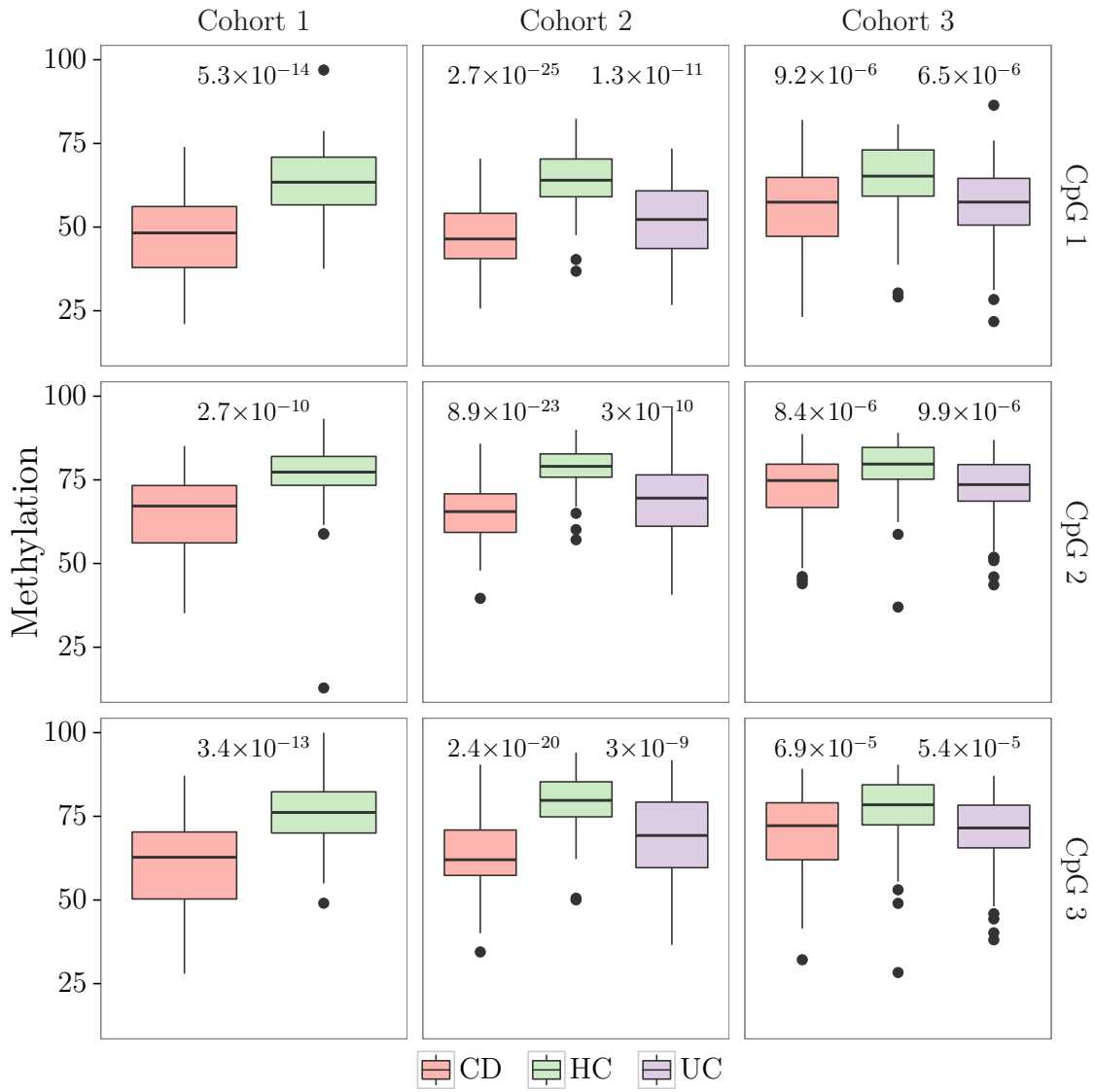


Figure 5.3: Box and whisker plots showing pyrosequencing results for all three adult cohorts at three CpGs (see table 5.3.1, page 215 for cohort and CpG details). Whiskers extend up to 1.5 × IQR from the 1st and 3rd quartiles, with any remaining results plotted as points. P values for Wilcoxon rank-sum tests between CD and control, and between UC and control.

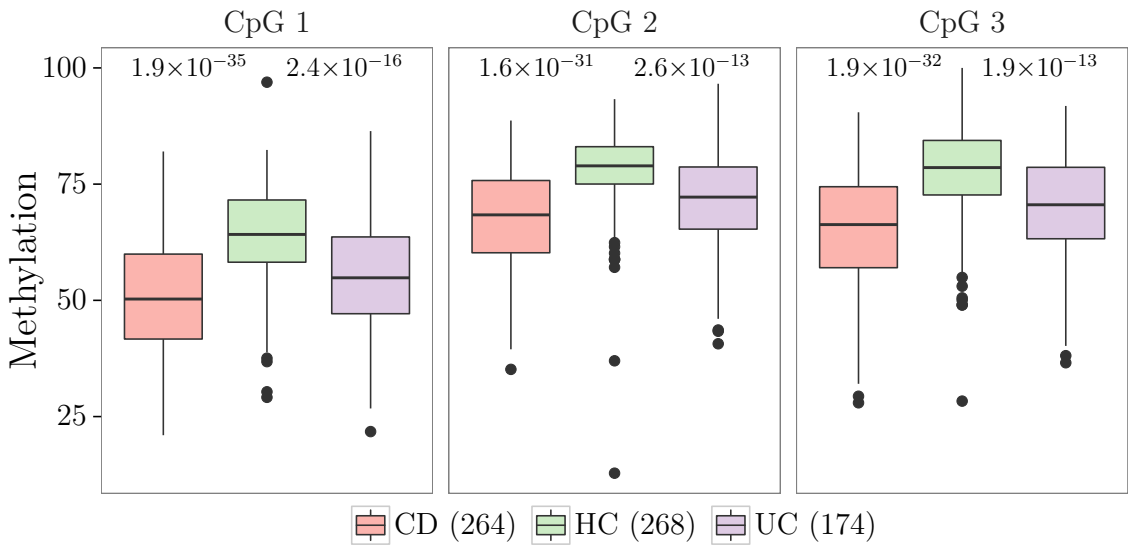


Figure 5.4: Box and whisker plots showing the combined adult pyrosequencing replication for RPS6KA2. Whiskers extend up to 1.5 × IQR from the 1st and 3rd quartiles, with any remaining results plotted as points. Individual cohort and CpG details given in table 5.4, page 215, combined numbers by diagnosis shown below figure.

5.3.2 Illumina 450k replication

Genome-wide methylation was assessed using the Illumina 450k platform in an adult cohort (largely overlapping with pyrosequencing cohort 2 in figure 5.3 and table 5.4) by work within our group undertaken by Ventham *et al.* [663][9] as part of the IBD-BIOM project. In this dataset the same *RPS6KA2* probe (cg17501210) was the most significant individual finding for Crohn's disease ($p=8.52\times10^{-25}$) and IBD ($p=6.85\times10^{-19}$), and ninth strongest for ulcerative colitis ($p=4.78\times10^{-12}$).

Analysis of the IBD Character Illumina 450k whole blood results (see appendix A.1) found the same *RPS6KA2* probe cg17501210 to have significant hypomethylation for CD, UC, and IBD overall, remaining significant if samples from Edinburgh were excluded. The β values are shown in figure 5.5 (page 219) and the p values in table 5.5 below.

	Rank	p value
IBD_{All}	3	1.08×10^{-19}
IBD_{Ed}	1	3.64×10^{-10}
IBD_{Other}	16	2.08×10^{-10}
CD	5	4.26×10^{-17}
UC	4	6.74×10^{-15}

Table 5.5: IBD Character Illumina 450k results for RPS6KA2 probe cg17501210

5.3.3 Biopsy microarray expression data

Analysis of microarray data[10] from inflamed and non-inflamed biopsies from patients with Crohn's disease or ulcerative colitis and controls shows an increase in *RPS6KA2* expression in inflamed Crohn's disease biopsies, not seen in inflamed biopsies from ulcerative colitis patients or controls, and no variation in baseline expression.

Expression levels of other RSK family members and closely related genes (listed in table 5.1, page 208), as well as *ERK1/2* and *p38* were also assessed. The only other transcript to show significant[11] disease- or inflammation-related expression changes was *RPS6KA6 (RSK4)*. Both significant sets of results are shown in figure 5.6 (page 220).

[9] Personal communication, results awaiting publication
[10] Experimental work: CLN
[11] corrected for multiple testing m=11

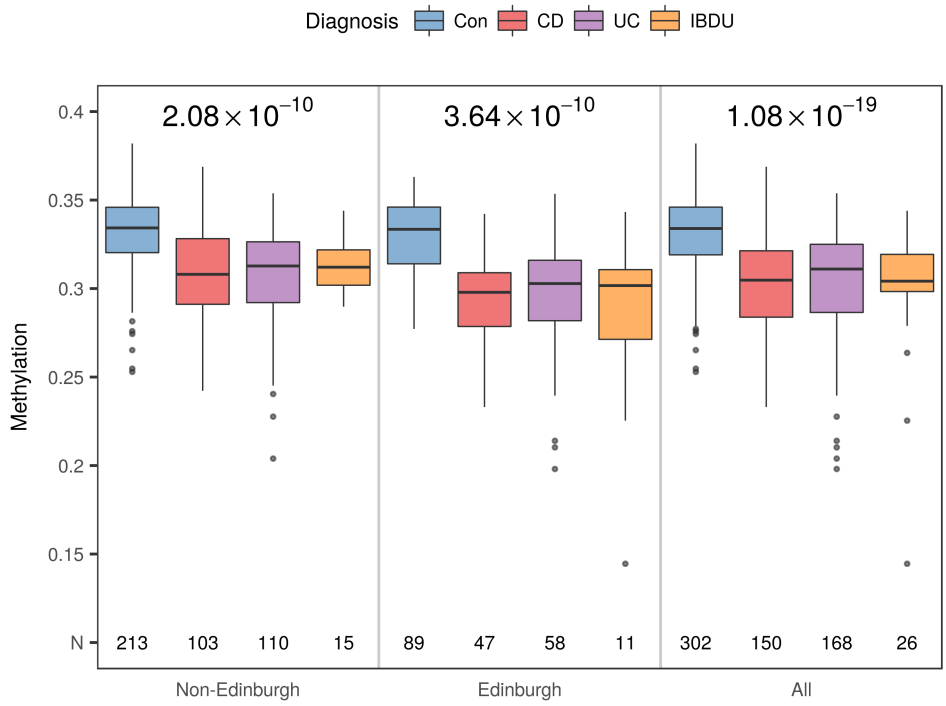


Figure 5.5: Box and whisker plots showing IBD Character RPS6KA2 Illumina 450k methylation results for all centres except Edinburgh (left), Edinburgh alone (centre), and all centres combined (right). Group sizes indicated below. Whiskers extend up to 1.5 × IQR from the 1st and 3rd quartiles, with any remaining results plotted as points. P values for limma analysis of IBD vs control for the two subgroup analyses and the overall analysis shown within each facet.

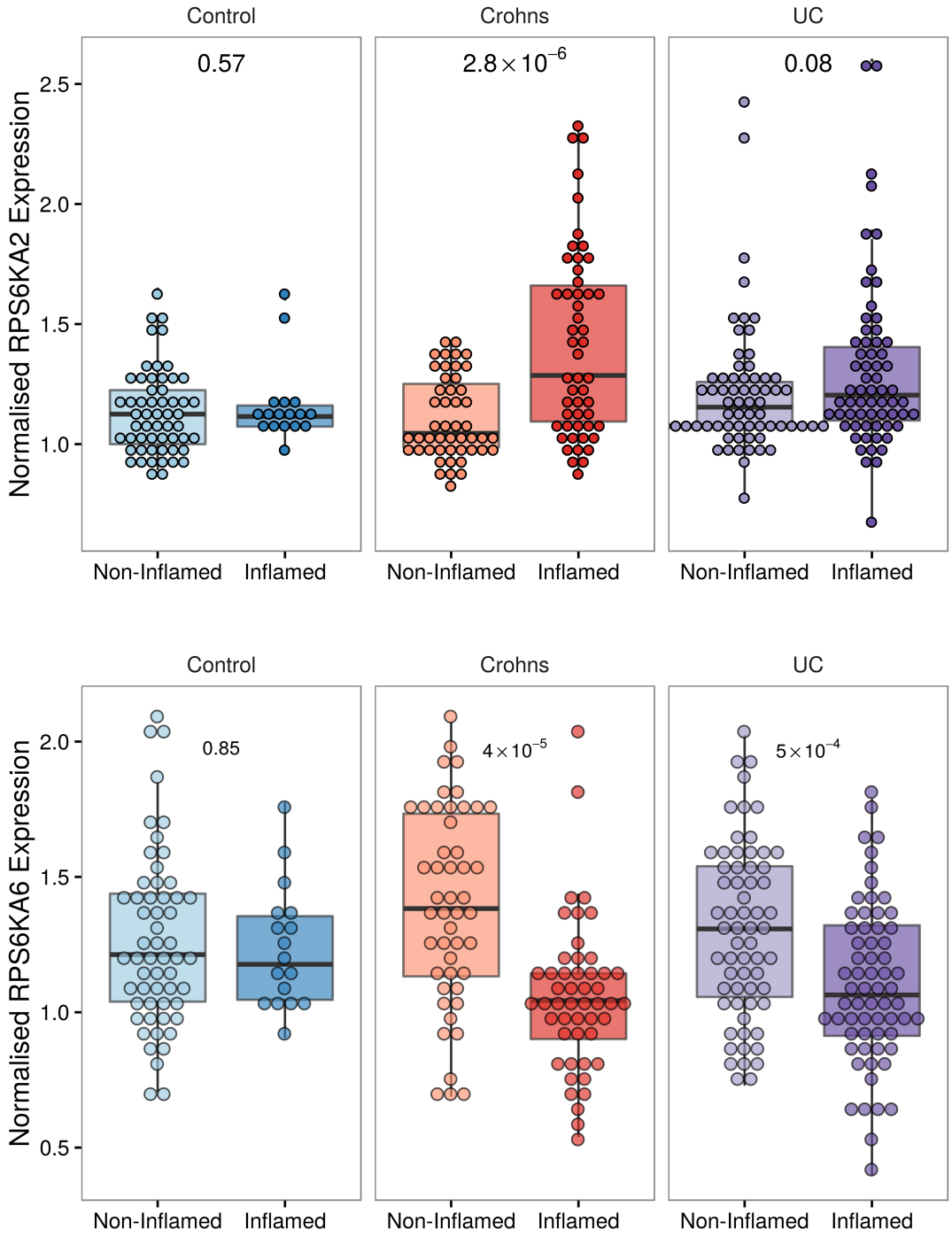


Figure 5.6: Box and whisker plots showing microarray data for *RSK3* (top) and *RSK4* (bottom) expression in non-inflamed and inflamed biopsies from controls (56 vs. 17), Crohn's disease (48 vs. 51), and ulcerative colitis (65 vs. 64). Whiskers extend up to 1.5 × IQR from the 1st and 3rd quartiles, with any remaining results plotted as points. P values for inflamed vs non-inflamed Wilcoxon rank sum test shown within each facet. Reanalysis of data from two papers by Noble *et al.* [461, 462], see section 3.2.1, page 52 for further details.

5.3.4 qPCR

No clear picture emerges from qPCR analysis of *RPS6KA2* mRNA expression levels (figure 5.7). There is a non-significant trend towards increased expression in whole blood, a non-significant trend towards decreased expression in PBMCs and no significant difference in expression between Crohn's disease and ulcerative colitis in granulocytes. No significant difference in the ratio of short and long transcripts was found between Crohn's disease and healthy controls.

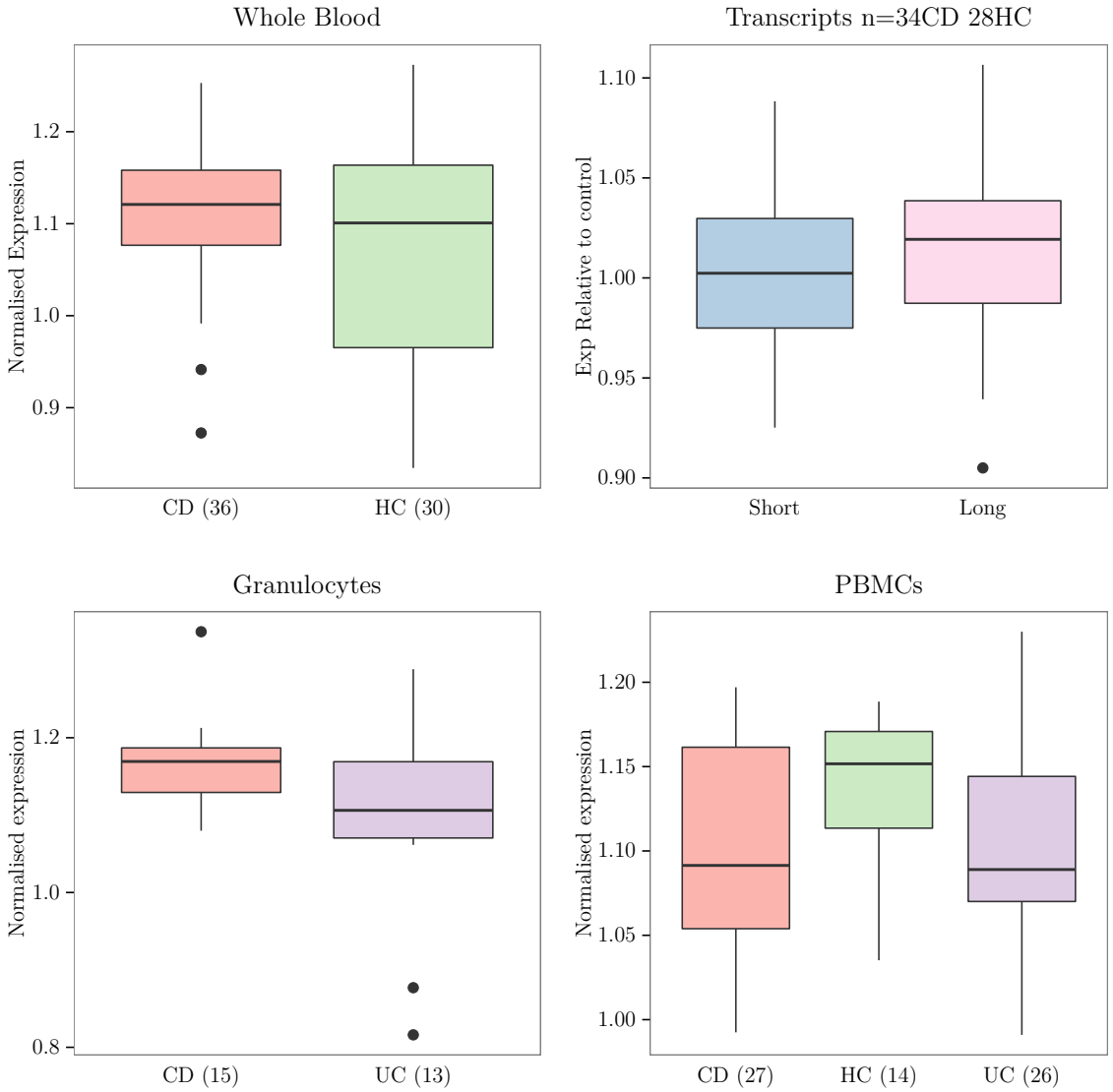


Figure 5.7: Box and whisker plots showing qPCR results for *RPS6KA2*. Whiskers extend up to 1.5 × IQR from the 1st and 3rd quartiles, with any remaining results plotted as points. Transcripts (top right) shows the relative expression of short and long transcripts in Crohn's disease normalised to median control expression. P values for all the shown comparisons were >0.05 by Wilcoxon rank sum test. Sample sizes shown in each facet.

5.3.5 Western blot

Knockdown of *RPS6KA2* was demonstrated by the first qPCR assay in table 5.3 (84% reduction) and by western blot. However, no significant difference in RPS6KA2 expression in response to stimulation of autophagy, and no significant difference in autophagy measured by LC3 western blot in response to RPS6KA2 knockdown were demonstrated.

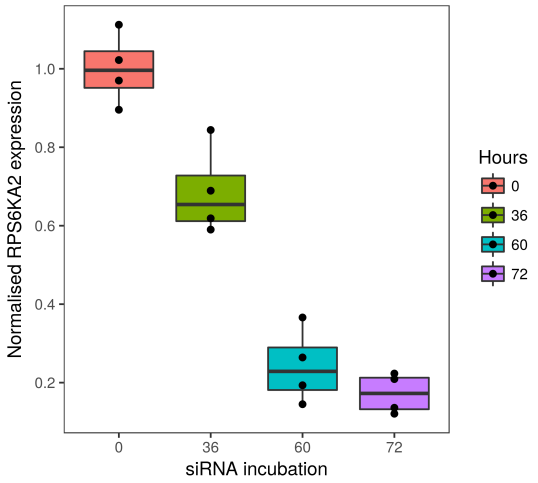


Figure 5.8: Expression of RPS6KA2 measured by qPCR after incubation with RPS6KA2 siRNA for 36–72 hours.

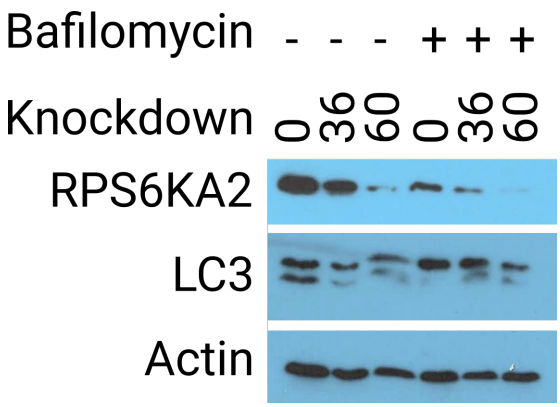


Figure 5.9: Western blot showing no significant effect of *RPS6KA2* knockdown (shown as hours exposure to RPS6KA2 siRNA) on autophagy

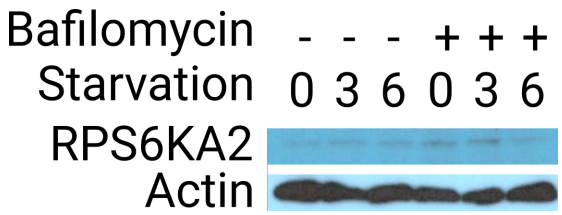


Figure 5.10: Western blot showing RPS6KA2 expression does not change with autophagy stimulation by starvation (0, 3, or 6 hours)

5.4 Discussion

Methylation Highly significant hypomethylation of the Illumina 450k probe cg17501210 in *RPS6KA2* has been shown in 1 paediatric cohort, and two large adult cohorts. In the IBD Character cohort the finding remained significant when all samples from Edinburgh were excluded, for a partially independent replication. Pyrosequencing of the same local region has also found highly significant hypomethylation in 3 adult cohorts (combined n=606, $p=1.9\times10^{-35}$).

This finding has not been reported in other IBD EWAS studies (table 2.4, page 33), though a weakly significant DMR was reported in an EWAS of treatment-naive paediatric UC biopsies [249]. However, as discussed in sections 2.5 and 4.4 comparisons between EWAS studies are complicated by differences in platforms and analytical techniques used, and tissues studied.

Al Muftah *et al.* found an association between probe cg17501210 and BMI ($p=4.90\times10^{-7}$)[12] in an EWAS for type 2 diabetes and BMI in a Qatari population [6]. This study showed other interesting overlaps with the paediatric Illumina 450k results; among their 8 replicated findings were also hypomethylation of *SOCS3* probe cg18181703 (the same probe showing Bonferroni significant hypomethylation in paediatric Crohn's disease), and hypomethylation of *SBNO2* probe cg07573872. Unfortunately this probe was not present in the paediatric Crohn's results, after failing QC, however the two closest probes (cg18608055 and cg12170787, 4.5 and 4.6kb distance) were both significantly hypomethylated after Bonferroni correction.

Expression Despite these highly replicable and significant methylation findings, little progress has been made in establishing any biological impact of the methylation change. While the microarray data (figure 5.6) showed increased expression in inflammation in CD, but not UC or controls, this was not replicated in the qPCR data (figure 5.7), and pre-publication data from a microarray study by NTV for two probes in *RPS6KA2* show weak changes in expression, with opposite directions[13].

The microarray probe presented here would measure all *RPS6KA2* transcripts 5.1, but qPCR primers designed to distinguish between the transcripts did not find any difference in the transcript ratio between CD and controls. Finally, no relationship between autophagy induced by starvation and RPS6KA2 expression

[12]Bonferroni corrected p=0.23 for m=468472
[13]Personal communication

was found with an antibody targeted to one of the regulatory phosphorylation sites in RPS6KA2.

There clearly is not a straightforward relationship between the methylation of this small intronic region of *RPS6KA2* and baseline expression. The nearby reported smoking-associated hypomethylation was also not found to correlate with expression [636]. The duration and timing of expression may be more important than baseline expression levels for RSK1-3 [449, 28] and further work should investigate this.

Genetics *RPS6KA2* lies within a GWAS risk locus for IBD [297] and genetic risk variants for colorectal cancer [580] and dental caries [729, 682] have also been described. As a solitary significant methylation probe, with no other nearby probes reaching even FDR significance, the possibility of a SNP must be considered. There are no SNPs within the probe as recognized by Illumina, though there are several of very low MAF, including the target base (figure 5.11, 227). There is only one probe with an appreciable MAF in the local region (global 0.34, European 0.156, British 0.165, table 5.6, page 228). Further work should include sequencing this region in a suitable number of patients and controls to establish if there is a SNP associated with IBD and whether it correlates with methylation of cg17501210.

Biological relevance Ribosomal protein S6 phosphorylation has been extensively studied, and multiple kinases have been shown phosphorylate the five conserved serine residues in the carboxy tail [62, 526]. Additionally drugs, pathological and physiological stimuli have been observed to have effects on RPS6 phosphorylation [62]. Despite this, the effects of RPS6 phosphorylation are incompletely understood, and while the p90 RSKs (including RPS6KA2) have shown RPS6 phosphorylation activity *in vitro* this may be mostly performed by the 70kDa S6 kinases *in vivo* [526], and other RPS6KA2 targets may be of much greater importance to IBD pathogenesis.

The MAPK/p38 pathway is important in signalling by TLRs [727], NOD1 [684], and NOD2 [254, 684]. Disruptions to p38 and ERK1/2 activity have been demonstrated in IBD and blockade has been shown to reduce the production of inflammatory cytokines including TNFα and IL-1β [673].

As discussed previously, *RPS6KA2* is in a GWAS risk locus for IBD [297] as are *RPS6KB1* [297], which shares a risk locus with the other strongest methylation finding *VMP1/MIR21*, *ERK2* [297], *p38a* and *ERk1* [673].

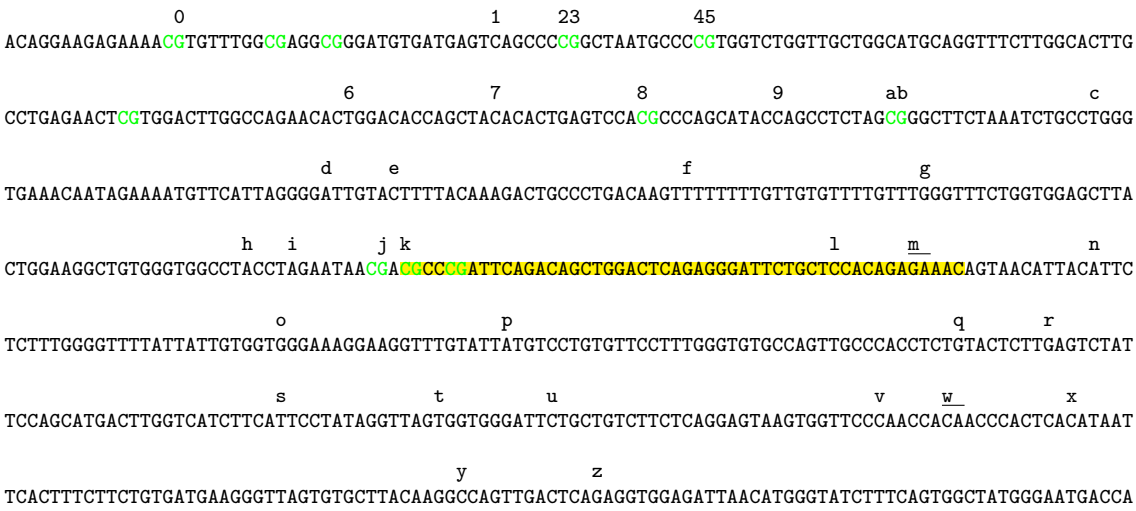


Figure 5.11: RPS6KA2 Illumina 450k probe cg17501210 with nearby SNPs indicated above (key shown in table 5.6, page 228), and CpGs highlighted in green

Key	SNP	Base	MAF	MAF_{EUR}	MAF_{GBR}	MAF_{TOP}
0	rs533103818	Y	-			
1	rs115081741	R	0.009	0	0	AFR 0.03, AMR 0.004
2	rs187153184	R	0.0002	0.01	0	CEU 0.05
3	rs552407586	Y	-			
4	rs182863394	R	0.0004	0	0	AFR 0.001, SAS 0.001
5	rs111294608	Y	0.014	0.001	0	AFR 0.05
6	rs190694832	M	0.003	0.001	0	AFR 0.007, AMR 0.07
7	rs536432002	R	0.001	0	0	AMR 0.007 (PUR 0.024)
8	rs762757999	R	-			
9	r767406436	R	-			
a	r73257221	K	0.058	0.002	0	AFR 0.21
b	r773795942	Y	-			
c	r553427214	W	0.0002	0	0	AFR 0.001
d	rs369745394	Y	-			
e	rs534709798	S	0.0004	0	0	AFR 0.002
f	rs768304238	R	-			
g	rs570783277	Y	0.0002	0	0	AFR 0.001
h	rs552609338	Y	0.001	0	0	EAS 0.007
i	rs748967106	Y	-			
j	rs537319626	Y	0.0004	0	0	AFR 0.002
k	rs569992380	R	0.0002	0	0	SAS 0.001
l	rs377062324	R	-			
m	rs748609690	TC/-	-			
n	rs771304997	Y	-			
o	rs77427546	Y	0.004	0	0	AFR 0.014
p	rs186180596	W	0.004	0.006	0.016	AMR 0.016
q	rs565685989	Y	0.0004	0	0	AFR 0.002
r	rs547424440	S	0.0002	0	0	AMR 0.001
s	rs182512563	R	0.0004	0	0	AFR 0.001
t	rs141041007	R	0.0001	0.001	0	CEU 0.005
u	rs9356490	K	-			
v	rs536851265	S	-			
w	rs575370430	TG/-	-			
x	rs758915478	K	-			
y	rs770201407	S	-			
z	rs971152	B	0.34	0.156	0.165	All

Table 5.6: SNPs near RPS6KA2 probe cg17501210 as shown in figure 5.11 by key. MAF: minor allele frequency overall, for Europeans (EUR), British in England and Scotland (GBR), and where most prevalent. Other population codes: African (AFR), Columbian, Mexican, Peruvian and Puerto Rican (AMR), east Asian (EAS), south Asian (SAS), Utah residents with northern and western European ancestry (CEU), Puerto Rican (PUR), no frequency data available (-).

Chapter 6

Biomarkers

6.1 Introduction

Diagnosis of IBD may be delayed in children, especially those without colonic involvement, and in those with non-specific symptoms [255]. If prolonged this can have profound effects on growth and education as discussed in section 2.6. Endoscopic investigation is the gold-standard investigation for the diagnosis of IBD, however a blood test would allow for a lower threshold for investigation which could reduce delayed diagnosis. A blood test would also be easier to repeat for monitoring purposes, would be cheaper, would not require access to endoscopy facilities and could therefore be carried out in general practice.

Epigenetic biomarkers are discussed in section 2.5.2 (page 36). DNA methylation can vary with disease, expression, genetic factors, and environmental exposure, making it a promising avenue for the discovery of biomarkers reflecting a combination of these factors.

The final paediatric methylation results presented in chapter 3 were the result of combining two paediatric datasets as discussed in section 3.3.1 (page 59). Primary analysis of the paediatric discovery cohort provided a small list of probes which later could be tested independently in the paediatric replication cohort.

The makeup of the paediatric discovery cohort is ideal for the development of diagnostic biomarkers. Both cases and controls had been referred for endoscopic investigation of symptoms suggestive of IBD, with DNA sampling and diagnosis occurring concurrently.

The work in this chapter was the basis for patent applications in the USA

[5][1] and internationally [3][2]. As part of these applications, biomarkers were also tested based on the top results in the combined paediatric dataset. Considering the performance of these in the paediatric patients is less scientifically rigorous, as they are constructed and tested against the same data; however, for comparisons against further datasets such as the adult IBD Character cohort is valid as there are once again independent learning and testing groups.

[1]Inventors: Adams, AT; Kennedy, NA; Nimmo, ER & Satsangi, J. 2014: 61/941847

[2]Inventors: Adams, AT; Ventham, NT; Kennedy, NA; Nimmo, ER & Satsangi, J. 2015: PCT/GB2015/050464

6.2 Materials & methods

Linear discriminant analysis of methylation beta values in the paediatric discovery cohort was used to create biomarkers for the presence of CD, using the LDA function in the R package 'MASS' [662]. All probes with FDR adjusted P values <0.05 in the initial analysis of the discovery cohort were used as covariates, regardless of performance in the replication cohort, with each model including 2 probes. Models were tested using the paediatric replication cohort methylation beta values.

For the patents applications, all 2080 combinations of two probes with Bonferroni significant methylation change in the combined paediatric analysis were tested. These models were built using the paediatric discovery cohort as a learning set and the paediatric replication cohort as a testing set.

Methylation results from the adult IBD Character cohort were used for a second replication of the paediatric models. For these analyses the entire paediatric cohort was used as a learning set. The IBD Character results were also used to test the 2080 models generated for the patent (excluding one probe which failed QC, leaving a total of 2016 combinations).

True and false, positive and negative results were determined from LDA, and sensitivity, specificity and accuracy calculated as shown below. Area under the curve (AUC) was calculated with the R package 'ROC' [96].

$$\text{Sens} = \frac{\sum \text{TP}}{\sum \text{TP} + \text{FN}} \qquad \text{Spec} = \frac{\sum \text{TN}}{\sum \text{TN} + \text{FP}} \qquad \text{Acc} = \frac{\sum \text{TP} + \text{TN}}{\sum \text{Samples}}$$

6.3 Results

6.3.1 Model generation & testing in children

Models were generated using the results of the 9 top probes in the discovery cohort (table 6.2, page 233). No single probe performed adequately as a biomarker, combining two probes was highly effective, but additional probes did not lead to significant improvements. Linear discriminant analysis of the 36 unique combinations of two significant probes in the paediatric discovery cohort (figure 6.1, page 234) was used to create models to predict diagnosis in the paediatric replication cohort. The equivalent methylation results in the replication cohort are shown in figure 6.2 (page 235), and the separation achieved in the replication cohort by these models, and the ROC curves are shown in figure 6.3 (page 236). The same models were constructed also including the age and sex as covariates, the results of which are shown in figure 6.4 (page 237).

The discovery cohort top results (table 6.2) include two *VMP1* probes, and each of these performed very similarly (as shown in figure 6.3), and *VMP1* is clearly involved in many of the better performing models. Therefore only one combination of each probe with a *VMP1* probe is included from this point to avoid overstating the overall accuracy of the models. The results of the application of each model to the paediatric replication cohort, including and excluding age and sex, are listed in table 6.3 (page 238), and table 6.4 (page 239) shows summary statistics for each group of models.

Overall, the inclusion of age and sex as covariates in the model improved sensitivity at the cost of reduced specificity; however, the differences are very small and a much larger cohort would be required to establish if inclusion of age and sex made a statistical difference to model performance. There was no advantage to adding further probes into models; the results of a model consisting of all 9 probes is shown in table 6.1.

Model	TP	FN	TN	FP	Sens	Spec	Acc	AUC
VMP1+MYO1E	14	3	18	0	0.82	1.00	0.91	0.98
9 Probes	13	4	17	1	0.76	0.94	0.86	0.95
VMP1+MYO1E+Age+Sex	15	2	18	0	0.88	0.89	0.94	0.96
9 Probes+Age+Sex	13	4	17	1	0.76	0.94	0.86	0.95

Table 6.1: The better two-probe models performed as well as using all 9 probes. T/F/P/N - true, false, positive, negative. Sens - sensitivity, Spec - specificity, Acc - accuracy, AUC - area under the curve.

				Discovery	Replication	Combined		
Probe	Chr	Coord	Gene	P value	P value	Rank	P value	Bonferroni
cg12054453	17	57915717	VMP1	2.19×10^{-10}	1.55×10^{-5}	1	1.97×10^{-15}	8.86×10^{-10}
cg16936953	17	57915665	VMP1	9.11×10^{-10}	9.25×10^{-5}	4	2.65×10^{-13}	1.19×10^{-7}
cg12992827	3	101901234		3.82×10^{-9}	5.14×10^{-5}	3	1.59×10^{-13}	7.13×10^{-8}
cg17501210	6	166970252	RPS6KA2	1.17×10^{-8}	3.73×10^{-6}	2	4.47×10^{-15}	2.01×10^{-9}
cg04389058	3	57041402	ARHGEF3	3.32×10^{-8}	4.69×10^{-5}	5	1.44×10^{-12}	6.48×10^{-7}
cg09349128	22	50327986		1.71×10^{-7}	3.46×10^{-3}	20	3.00×10^{-9}	1.35×10^{-3}
cg25114611	6	35696870		4.72×10^{-7}	0.021	102	3.05×10^{-7}	0.137
cg08423142	15	59588622	MYO1E	4.86×10^{-7}	0.017	25	4.07×10^{-9}	1.83×10^{-3}
cg22628873	17	4464400	GGT6	7.55×10^{-7}	0.101	136	6.90×10^{-7}	0.310

Table 6.2: Probes selected for use in biomarkers, based on their significance in the discovery cohort regardless of the significance in the replication cohort. Also shown are the rank, p value, and Bonferroni corrected p value in the combined analysis.

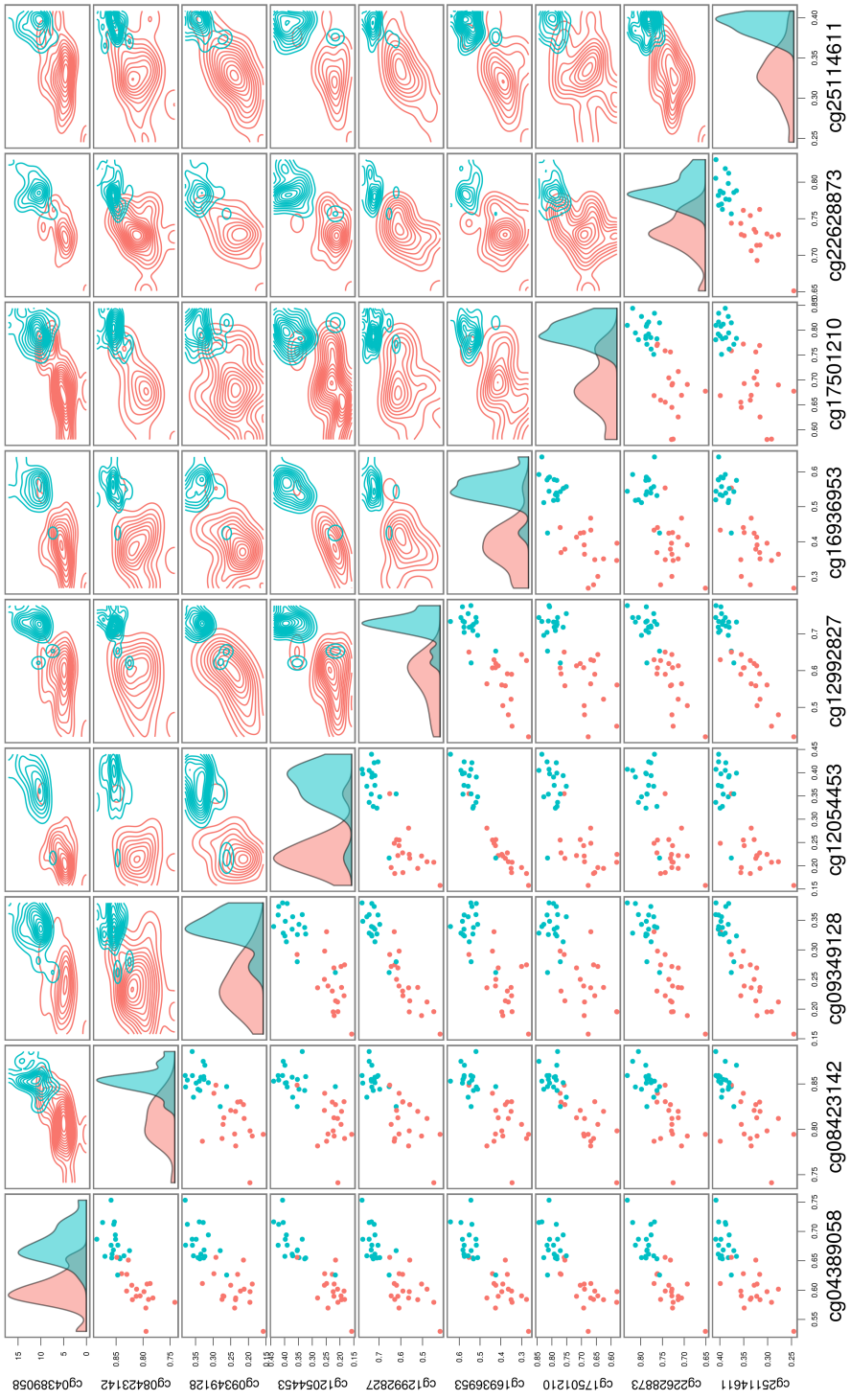


Figure 6.1: Results in the paediatric discovery cohort for all two-probe combinations used for training diagnostic biomarkers; lower-left - individual β values for each two-probe combination; upper-right - 2D probability density by diagnosis; diagonal - 1D probability density. Red: CD, blue: control.

Figure 6.2: Results in the paediatric replication cohort for top probes in the paediatric discovery cohort used to train models; lower-left - individual β values for each two-probe combination; upper-right - 2D probability density by diagnosis; diagonal - 1D probability density. Red: CD, blue: control.

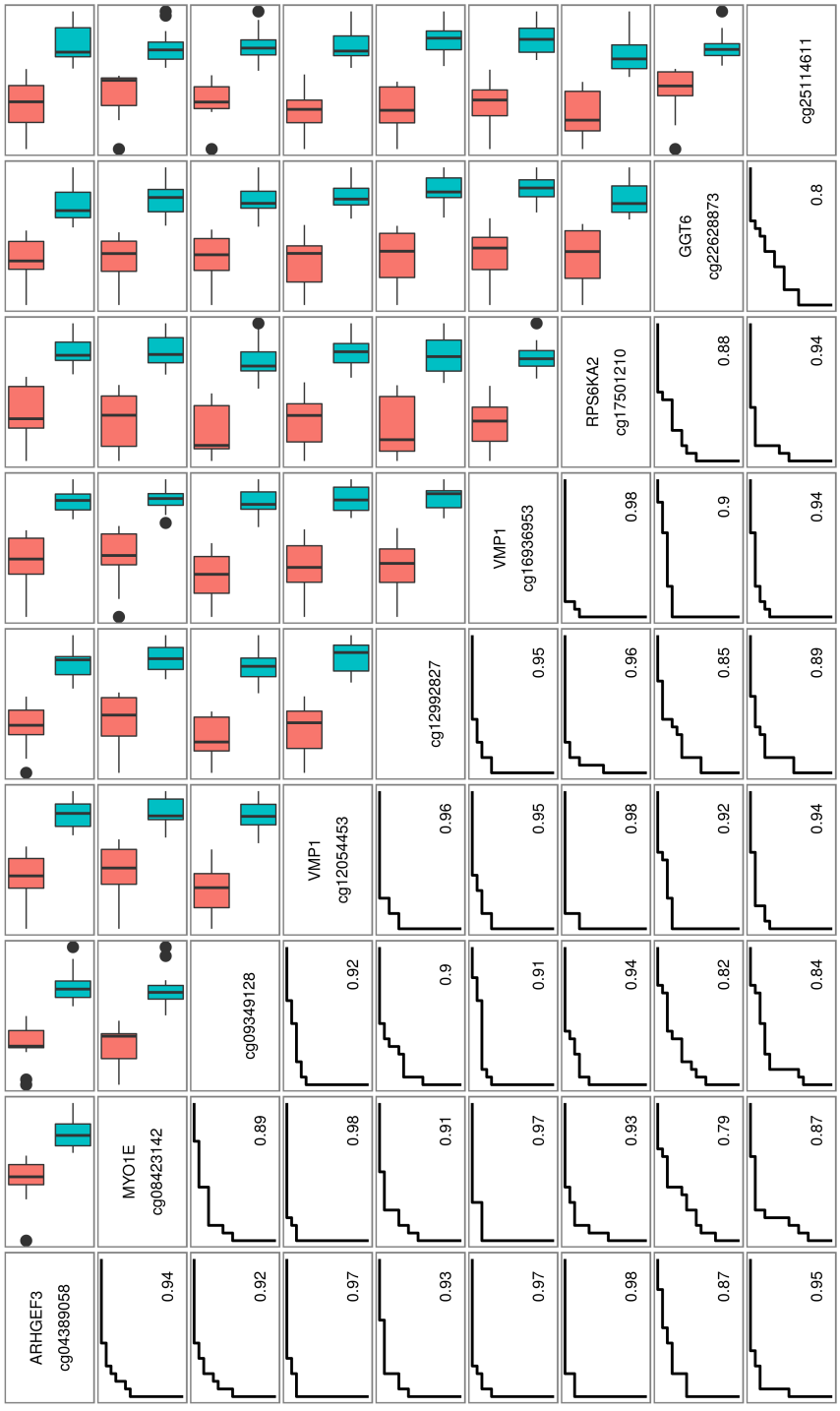


Figure 6.3: Upper-right - box and whisker plots showing separation in the replication cohort by the linear discriminant function derived from the testing cohort. Whiskers extend up to 1.5 $\times$ IQR from the 1st and 3rd quartiles, with any remaining results plotted as points; lower-left - ROC plots and AUC for the performance of these models in the replication cohort. Red: CD, blue: control.

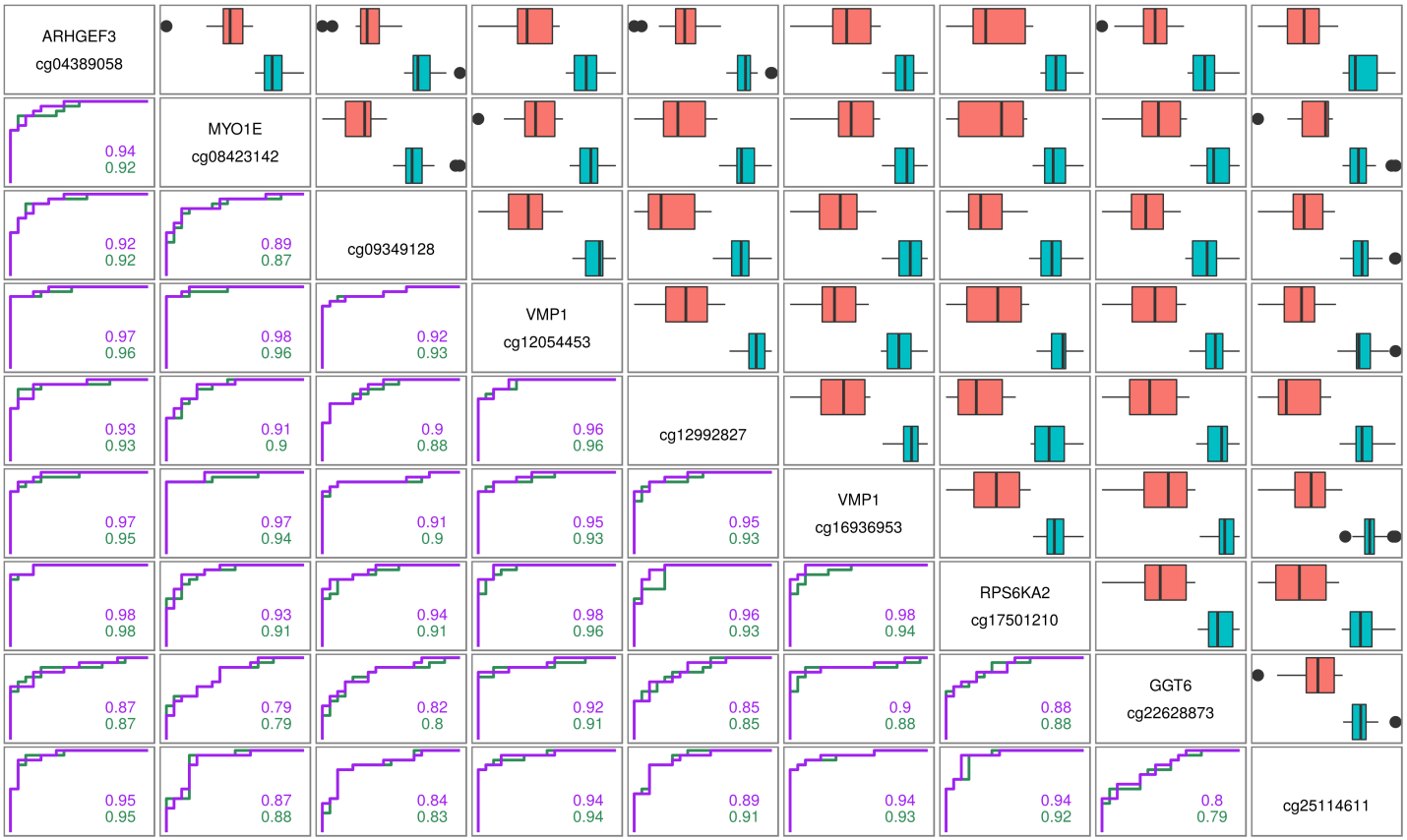


Figure 6.4: Upper-right - box and whisker plots showing separation in the replication cohort by the linear discriminant function derived from the testing cohort, including age and sex as covariates. Whiskers extend up to 1.5 × IQR from the 1st and 3rd quartiles, with any remaining results plotted as points; lower-left - ROC plots and AUC for the performance in the replication cohort of models previously shown, or including subject age and sex. Red: CD, blue: control.

Probes		Methylation								Methylation, age & sex							
		TP	FN	TN	FP	Sens	Spec	Acc	AUC	TP	FN	TN	FP	Sens	Spec	Acc	AUC
VMP1	MYO1E	14	3	18	0	0.82	1.00	0.91	0.98	15	2	18	0	0.88	1.00	0.94	0.96
VMP1	RPS6KA2	13	4	18	0	0.76	1.00	0.89	0.98	14	3	16	2	0.82	0.89	0.86	0.96
RPS6KA2	ARHGEF3	13	4	18	0	0.76	1.00	0.89	0.98	14	3	17	1	0.82	0.94	0.89	0.98
VMP1	ARHGEF3	13	4	18	0	0.76	1.00	0.89	0.97	13	4	18	0	0.76	1.00	0.89	0.96
VMP1	cg12992827	13	4	18	0	0.76	1.00	0.89	0.96	14	3	17	1	0.82	0.94	0.89	0.96
RPS6KA2	cg12992827	12	5	17	1	0.71	0.94	0.83	0.96	11	6	17	1	0.65	0.94	0.80	0.93
VMP1	VMP1	13	4	16	2	0.76	0.89	0.83	0.95	13	4	17	1	0.76	0.94	0.86	0.93
ARHGEF3	cg25114611	13	4	17	1	0.76	0.94	0.86	0.95	13	4	17	1	0.76	0.94	0.86	0.95
VMP1	cg25114611	13	4	17	1	0.76	0.94	0.86	0.94	14	3	17	1	0.82	0.94	0.89	0.94
RPS6KA2	cg09349128	12	5	17	1	0.71	0.94	0.83	0.94	12	5	16	2	0.71	0.89	0.80	0.91
RPS6KA2	cg25114611	11	6	17	1	0.65	0.94	0.80	0.94	11	6	15	3	0.65	0.83	0.74	0.92
ARHGEF3	MYO1E	12	5	16	2	0.71	0.89	0.80	0.94	13	4	17	1	0.76	0.94	0.86	0.92
RPS6KA2	MYO1E	11	6	17	1	0.65	0.94	0.80	0.93	13	4	16	2	0.76	0.89	0.83	0.91
cg12992827	ARHGEF3	12	5	17	1	0.71	0.94	0.83	0.93	13	4	17	1	0.76	0.94	0.86	0.93
VMP1	cg09349128	13	4	17	1	0.76	0.94	0.86	0.92	14	3	16	2	0.82	0.89	0.86	0.93
VMP1	GGT6	14	3	18	0	0.82	1.00	0.91	0.92	13	4	16	2	0.76	0.89	0.83	0.91
ARHGEF3	cg09349128	12	5	17	1	0.71	0.94	0.83	0.92	12	5	16	2	0.71	0.89	0.80	0.92
cg12992827	MYO1E	13	4	14	4	0.76	0.78	0.77	0.91	12	5	16	2	0.71	0.89	0.80	0.90
cg12992827	cg09349128	12	5	16	2	0.71	0.89	0.80	0.90	12	5	15	3	0.71	0.83	0.77	0.88
cg12992827	cg25114611	12	5	16	2	0.71	0.89	0.80	0.89	12	5	16	2	0.71	0.89	0.80	0.91
cg09349128	MYO1E	14	3	15	3	0.82	0.83	0.83	0.89	13	4	16	2	0.76	0.89	0.83	0.87
RPS6KA2	GGT6	11	6	16	2	0.65	0.89	0.77	0.88	11	6	15	3	0.65	0.83	0.74	0.88
ARHGEF3	GGT6	11	6	16	2	0.65	0.89	0.77	0.87	12	5	16	2	0.71	0.89	0.80	0.87
MYO1E	cg25114611	10	7	15	3	0.59	0.83	0.71	0.87	10	7	15	3	0.59	0.83	0.71	0.88
cg12992827	GGT6	12	5	14	4	0.71	0.78	0.74	0.85	12	5	15	3	0.71	0.83	0.77	0.85
cg09349128	cg25114611	13	4	15	3	0.76	0.83	0.80	0.84	13	4	15	3	0.76	0.83	0.80	0.83
cg09349128	GGT6	12	5	14	4	0.71	0.78	0.74	0.82	12	5	15	3	0.71	0.83	0.77	0.80
cg25114611	GGT6	11	6	13	5	0.65	0.72	0.69	0.80	10	7	13	5	0.59	0.72	0.66	0.79
MYO1E	GGT6	9	8	14	4	0.53	0.78	0.66	0.79	11	6	14	4	0.65	0.78	0.71	0.79

Table 6.3: Two-probe biomarkers, as shown in figures 6.3 (left) & 6.4 (right), excluding the similar results from substituting between the two *VMP1* probes). T/F/P/N - true, false, positive, negative. Sens - sensitivity, Spec - specificity, Acc - accuracy, AUC - area under the curve.

	Methylation			Meth, age & sex		
	Min	Median	Max	Min	Median	Max
Sensitivity	0.53	0.71	0.82	0.59	0.76	0.88
Specificity	0.72	0.94	1.00	0.72	0.89	1.00
Accuracy	0.66	0.83	0.91	0.66	0.80	0.94
AUC	0.79	0.92	0.98	0.79	0.91	0.98

Table 6.4: Summary results for paediatric biomarkers

6.3.2 Testing of paediatric models in adults

Many of the probes used in the previous section were also highly significantly differentially methylated in the adult IBD Character cohort (see table 6.5). Overall there was a greater overlap between the methylation values in Crohn's and control in adults compared to children (see figure 6.5, page 241). Additionally, in probes which were highly significant in both the children and adults, with the same direction and magnitude of change in methylation, the absolute methylation values did not always match between children and adults (see figure 6.6, page 242). The combined effect of these factors is a reduced accuracy of the models when tested in adults, as shown in figure 6.7 (page 243), and table 6.6 (page 244).

Probe	Chr	Coord	Gene	P value
cg04389058	3	57041402	ARHGEF3	7.14×10^{-8}
cg08423142	15	59588622	MYO1E	1.49×10^{-3}
cg09349128	22	50327986	CRELD2	8.41×10^{-21}
cg12054453	17	57915717	VMP1	3.66×10^{-20}
cg12992827	3	101901234	ZPLD1	2.60×10^{-11}
cg16936953	17	57915665	VMP1	4.05×10^{-18}
cg17501210	6	166970252	RPS6KA2	1.08×10^{-19}
cg22628873	17	4464400	GGT6	1.28×10^{-3}
cg25114611	6	35696870	LOC285847	1.27×10^{-8}

Table 6.5: Illumina 450k results in the IBD Character dataset for the paediatric biomarker probes

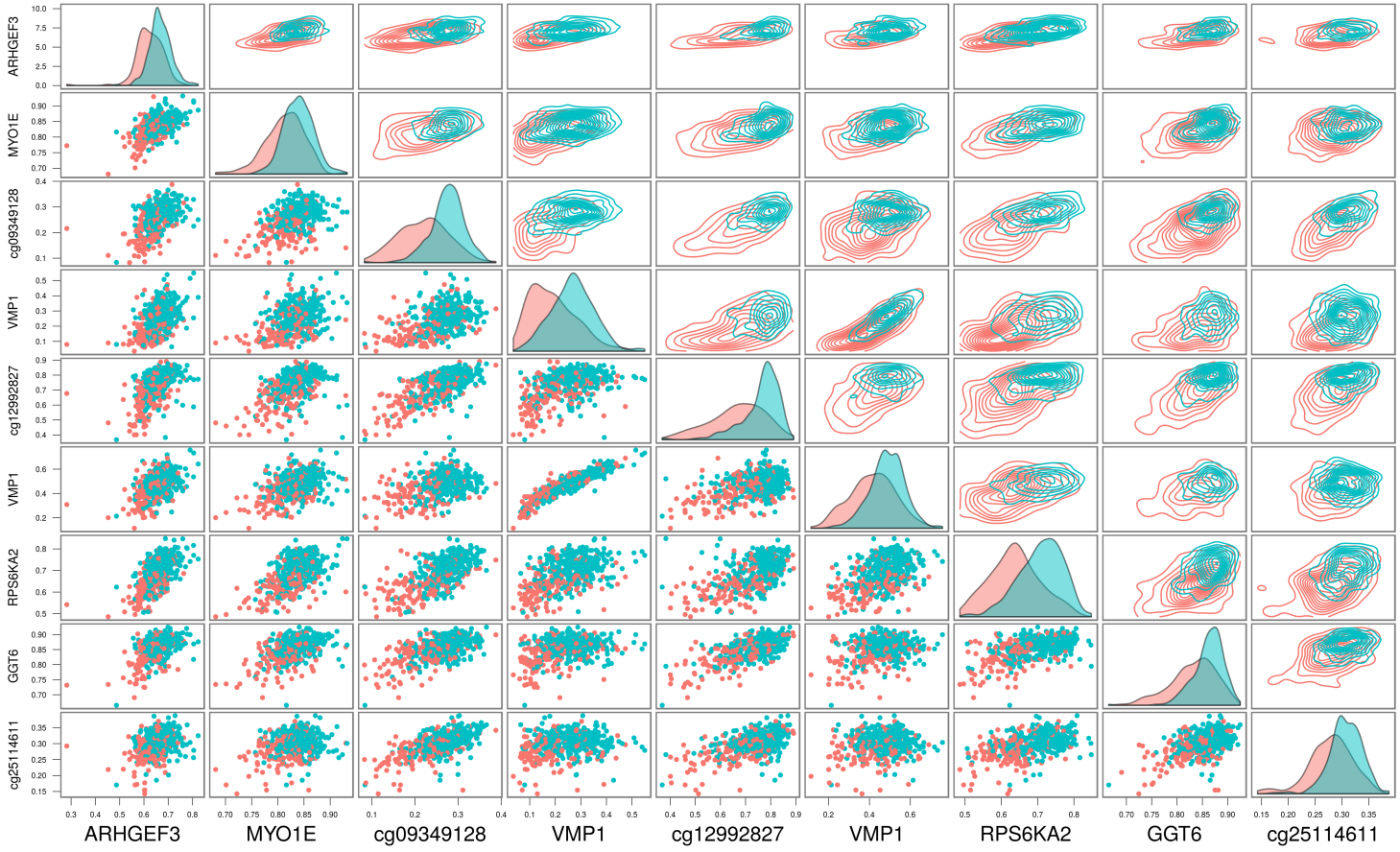


Figure 6.5: Results in the adult IBD Character dataset for top probes in the paediatric discovery cohort used to train models; lower-left - individual β values for each two-probe combination; upper-right - 2D probability density by diagnosis; diagonal - 1D probability density. Red: CD, blue: control.

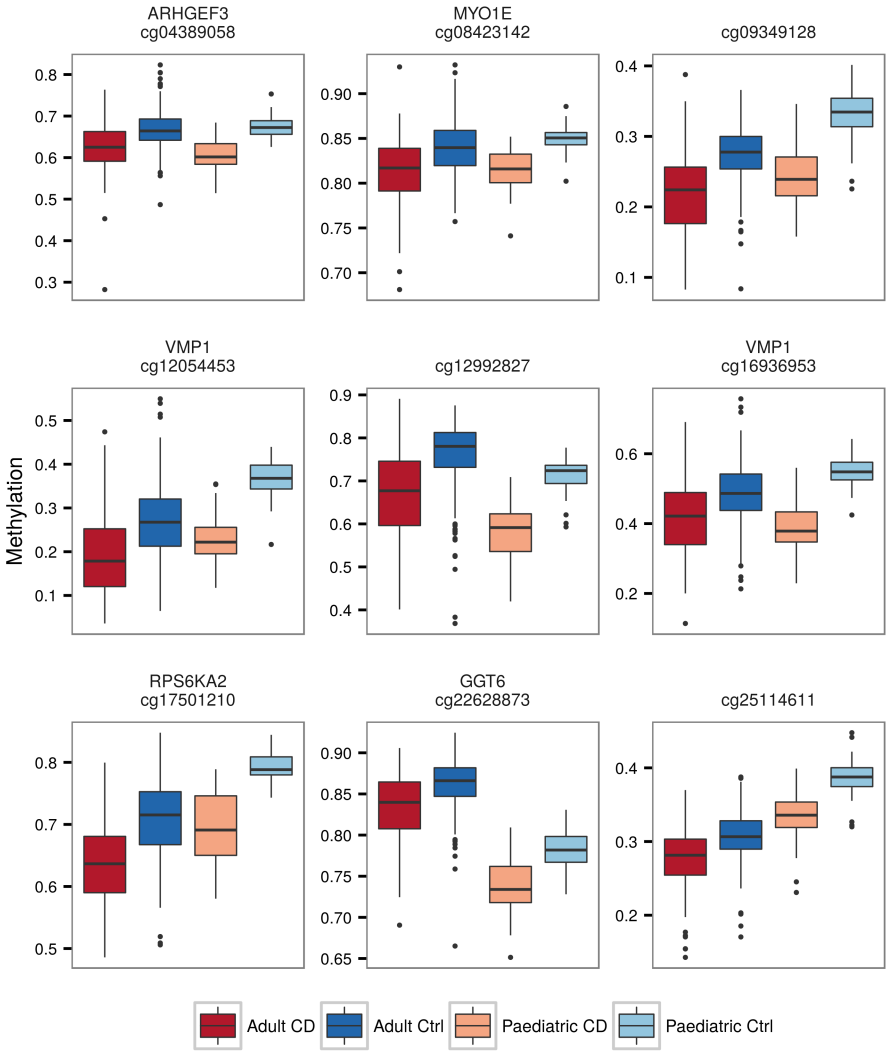


Figure 6.6: Box and whisker plots showing methylation results in adults and children, for probes which were highly significant in both analyses. Whiskers extend up to 1.5 $\times$ IQR from the 1st and 3rd quartiles, with any remaining results plotted as points.

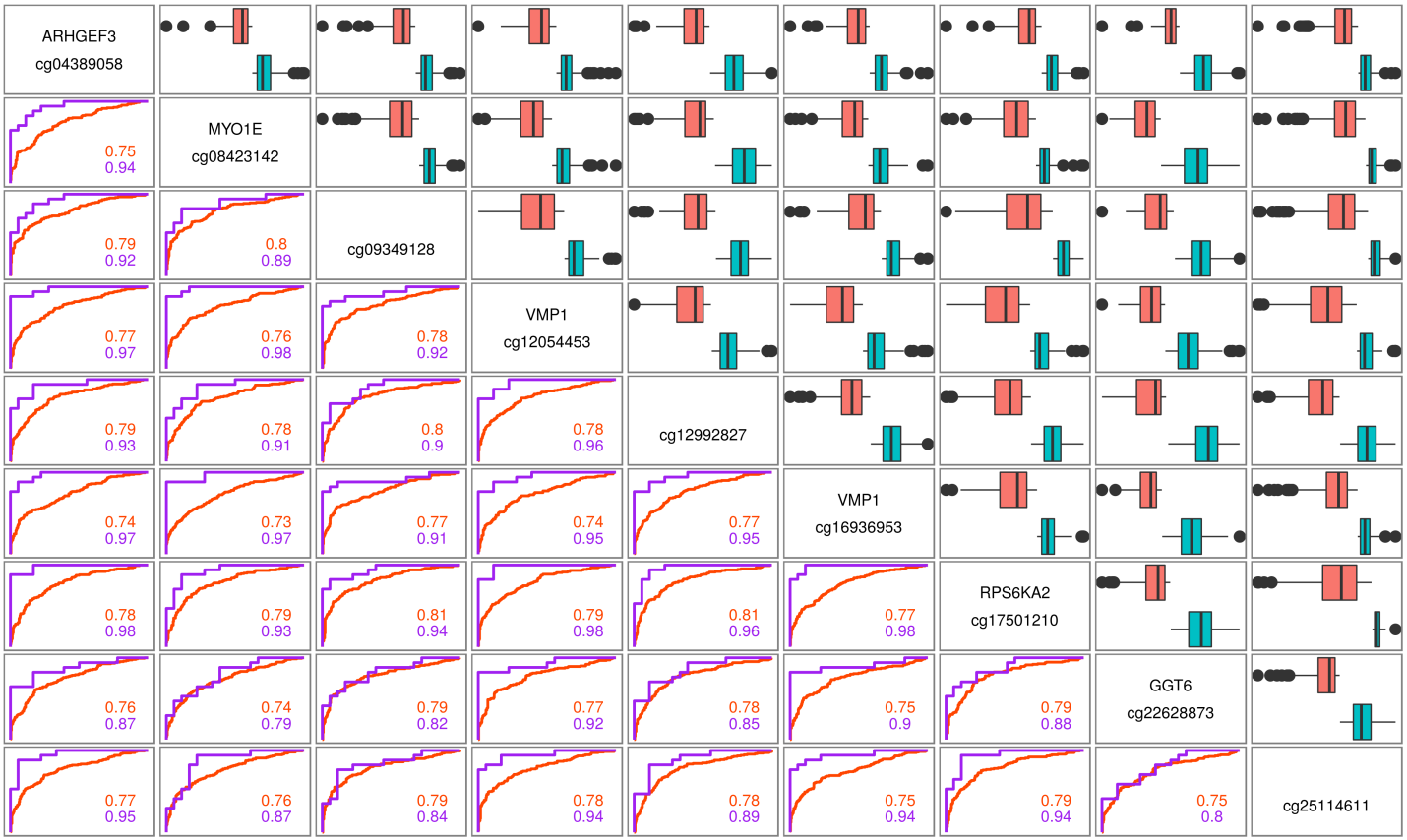


Figure 6.7: Upper-right - box and whisker plots showing separation in the adult cohort by the linear discriminant function derived from the paediatric testing cohort. Whiskers extend up to 1.5 $\times$ IQR from the 1st and 3rd quartiles, with any remaining results plotted as points; lower-left - ROC plots and AUC for the performance in the adult cohort compared to paediatric testing cohort. Red: CD, blue: control.

Probes		Methylation								Methylation, age & sex							
		TP	FN	TN	FP	Sens	Spec	Acc	AUC	TP	FN	TN	FP	Sens	Spec	Acc	AUC
cg12992827	ARHGEF3	71	79	269	33	0.47	0.89	0.75	0.79	41	109	293	9	0.27	0.97	0.74	0.79
ARHGEF3	cg09349128	113	37	188	114	0.75	0.62	0.67	0.79	86	64	275	27	0.57	0.91	0.80	0.79
cg09349128	MYO1E	128	22	163	139	0.85	0.54	0.64	0.80	90	60	256	46	0.60	0.85	0.77	0.78
VMP1	cg25114611	146	4	52	250	0.97	0.17	0.44	0.78	130	20	110	192	0.87	0.36	0.53	0.78
ARHGEF3	cg25114611	128	22	110	192	0.85	0.36	0.53	0.77	113	37	198	104	0.75	0.66	0.69	0.79
VMP1	RPS6KA2	140	10	91	211	0.93	0.30	0.51	0.79	95	55	241	61	0.63	0.80	0.74	0.76
VMP1	cg12992827	109	41	215	87	0.73	0.71	0.72	0.78	86	64	269	33	0.57	0.89	0.79	0.77
VMP1	cg09349128	134	16	101	201	0.89	0.33	0.52	0.78	118	32	180	122	0.79	0.60	0.66	0.77
RPS6KA2	cg12992827	104	46	237	65	0.69	0.78	0.75	0.81	48	102	294	8	0.32	0.97	0.76	0.74
RPS6KA2	ARHGEF3	125	25	168	134	0.83	0.56	0.65	0.78	63	87	286	16	0.42	0.95	0.77	0.77
cg12992827	cg09349128	79	71	265	37	0.53	0.88	0.76	0.80	43	107	291	11	0.29	0.96	0.74	0.75
VMP1	ARHGEF3	126	24	141	161	0.84	0.47	0.59	0.77	121	29	172	130	0.81	0.57	0.65	0.77
VMP1	GGT6	88	62	261	41	0.59	0.86	0.77	0.77	62	88	280	22	0.41	0.93	0.76	0.77
RPS6KA2	cg09349128	141	9	97	205	0.94	0.32	0.53	0.81	74	76	275	27	0.49	0.91	0.77	0.73
cg12992827	MYO1E	70	80	269	33	0.47	0.89	0.75	0.78	43	107	290	12	0.29	0.96	0.74	0.76
cg09349128	cg25114611	142	8	44	258	0.95	0.15	0.41	0.79	100	50	207	95	0.67	0.69	0.68	0.75
VMP1	MYO1E	125	25	131	171	0.83	0.43	0.57	0.76	122	28	157	145	0.81	0.52	0.62	0.77
RPS6KA2	cg25114611	146	4	21	281	0.97	0.07	0.37	0.79	86	64	239	63	0.57	0.79	0.72	0.74
ARHGEF3	GGT6	34	116	296	6	0.23	0.98	0.73	0.76	22	128	300	2	0.15	0.99	0.71	0.77
cg09349128	GGT6	52	98	289	13	0.35	0.96	0.75	0.79	28	122	300	2	0.19	0.99	0.73	0.74
RPS6KA2	MYO1E	133	17	138	164	0.89	0.46	0.60	0.79	63	87	280	22	0.42	0.93	0.76	0.73
RPS6KA2	GGT6	63	87	275	27	0.42	0.91	0.75	0.79	19	131	300	2	0.13	0.99	0.71	0.73
cg12992827	VMP1	74	76	267	35	0.49	0.88	0.75	0.77	49	101	289	13	0.33	0.96	0.75	0.75
cg12992827	cg25114611	74	76	266	36	0.49	0.88	0.75	0.78	37	113	292	10	0.25	0.97	0.73	0.74
cg12992827	GGT6	32	118	295	7	0.21	0.98	0.72	0.78	14	136	299	3	0.09	0.99	0.69	0.74
MYO1E	cg25114611	142	8	58	244	0.95	0.19	0.44	0.76	126	24	137	165	0.84	0.45	0.58	0.76
VMP1	cg09349128	116	34	167	135	0.77	0.55	0.63	0.77	80	70	267	35	0.53	0.88	0.77	0.74
ARHGEF3	MYO1E	95	55	227	75	0.63	0.75	0.71	0.75	82	68	249	53	0.55	0.82	0.73	0.76
RPS6KA2	VMP1	127	23	152	150	0.85	0.50	0.62	0.77	72	78	278	24	0.48	0.92	0.77	0.73
VMP1	ARHGEF3	92	58	214	88	0.61	0.71	0.68	0.74	74	76	262	40	0.49	0.87	0.74	0.76
VMP1	cg25114611	134	16	98	204	0.89	0.32	0.51	0.75	98	52	207	95	0.65	0.69	0.67	0.74
VMP1	GGT6	39	111	291	11	0.26	0.96	0.73	0.75	23	127	299	3	0.15	0.99	0.71	0.74
VMP1	VMP1	124	26	121	181	0.83	0.40	0.54	0.74	114	36	165	137	0.76	0.55	0.62	0.74
MYO1E	GGT6	27	123	295	7	0.18	0.98	0.71	0.74	17	133	300	2	0.11	0.99	0.70	0.74
cg25114611	GGT6	60	90	275	27	0.40	0.91	0.74	0.75	31	119	300	2	0.21	0.99	0.73	0.73
VMP1	MYO1E	102	48	199	103	0.68	0.66	0.67	0.73	78	72	249	53	0.52	0.82	0.72	0.73

Table 6.6: Paediatric models tested in the IBD Character adult cohort. T/F/P/N - true, false, positive, negative. Sens - sensitivity, Spec - specificity, Acc - accuracy, AUC - area under the curve.

6.3.3 Further paediatric modelling

For the patent applications, all 2080 combinations of the 65 Bonferroni significant probes were tested. The top results in children using the discovery cohort as the learning set and the replication cohort as the testing set are shown in table 6.7 (page 246).

As with the previously discussed probes (figure 6.6, page 242) there were frequently substantial shifts in the methylation patterns seen between children and adults, even when the probe was highly significant in both cohorts, with a similar magnitude of disease-associated change in the same direction. This is shown in figure 6.8 (page 247), showing a generally symmetrical shift in methylation for Crohn's and controls, and a preponderance of hypomethylation in adults relative to children. One of the 65 Bonferroni significant paediatric results was filtered during the adult data QC, leaving 64 probes and 2016 two-probe combinations.

As with the previous models, these models, constructed in children, performed less well in adults than children. The best performing of these models in adults, using the entire paediatric cohort as a learning set, are shown in table 6.8 (page 248), and the performance of all 2016 models are shown in figure 6.9 (page 249).

Probes		Methylation								Methylation, age & sex							
		TP	FN	TN	FP	Sens	Spec	Acc	AUC	TP	FN	TN	FP	Sens	Spec	Acc	AUC
RPS6KA2	HK2	16	1	18	0	0.94	1.00	0.97	1.00	16	1	18	0	0.94	1.00	0.97	1.00
RPS6KA2	cg01726890	15	2	18	0	0.88	1.00	0.94	1.00	15	2	18	0	0.88	1.00	0.94	1.00
RPS6KA2	ZBTB16	17	0	18	0	1.00	1.00	1.00	1.00	17	0	18	0	1.00	1.00	1.00	1.00
RPS6KA2	ZEB2	12	5	18	0	0.71	1.00	0.86	1.00	12	5	18	0	0.71	1.00	0.86	1.00
RPS6KA2	CDC42BPB	12	5	18	0	0.71	1.00	0.86	1.00	12	5	18	0	0.71	1.00	0.86	1.00
RPS6KA2	SLC15A4	17	0	18	0	1.00	1.00	1.00	1.00	17	0	18	0	1.00	1.00	1.00	1.00
HK2	TNFSF10	15	2	18	0	0.88	1.00	0.94	1.00	15	2	18	0	0.88	1.00	0.94	1.00
HK2	FKBP5	16	1	18	0	0.94	1.00	0.97	1.00	16	1	18	0	0.94	1.00	0.97	1.00
TNFSF10	IL18RAP	17	0	17	1	1.00	0.94	0.97	1.00	17	0	17	1	1.00	0.94	0.97	1.00
SOCS3	IL18RAP	17	0	17	1	1.00	0.94	0.97	1.00	17	0	17	1	1.00	0.94	0.97	1.00
cg01726890	SBNO2	17	0	17	1	1.00	0.94	0.97	1.00	17	0	17	1	1.00	0.94	0.97	1.00
cg01726890	FKBP5	17	0	17	1	1.00	0.94	0.97	1.00	17	0	17	1	1.00	0.94	0.97	1.00
ZBTB16	SBNO2	17	0	18	0	1.00	1.00	1.00	1.00	17	0	18	0	1.00	1.00	1.00	1.00
ZBTB16	FKBP5	17	0	18	0	1.00	1.00	1.00	1.00	17	0	18	0	1.00	1.00	1.00	1.00
SBNO2	SLC15A4	16	1	18	0	0.94	1.00	0.97	1.00	16	1	18	0	0.94	1.00	0.97	1.00
IL18RAP	SBNO2	17	0	18	0	1.00	1.00	1.00	1.00	17	0	18	0	1.00	1.00	1.00	1.00
IL18RAP	MIR21	16	1	18	0	0.94	1.00	0.97	1.00	16	1	18	0	0.94	1.00	0.97	1.00
IL18RAP	FKBP5	17	0	16	2	1.00	0.89	0.94	1.00	17	0	16	2	1.00	0.89	0.94	1.00
VMP1	SLC15A4	14	3	18	0	0.82	1.00	0.91	1.00	14	3	18	0	0.82	1.00	0.91	1.00
RPS6KA2	IL18RAP	16	1	18	0	0.94	1.00	0.97	1.00	16	1	18	0	0.94	1.00	0.97	1.00
HK2	SOCS3	16	1	17	1	0.94	0.94	0.94	1.00	16	1	17	1	0.94	0.94	0.94	1.00
SOCS3	ZBTB16	17	0	17	1	1.00	0.94	0.97	1.00	17	0	17	1	1.00	0.94	0.97	1.00
VMP1	ZBTB16	16	1	18	0	0.94	1.00	0.97	1.00	16	1	18	0	0.94	1.00	0.97	1.00
SBNO2	CBFA2T2	15	2	18	0	0.88	1.00	0.94	1.00	15	2	18	0	0.88	1.00	0.94	1.00
VMP1	IL18RAP	15	2	18	0	0.88	1.00	0.94	0.99	15	2	18	0	0.88	1.00	0.94	0.99
RPS6KA2	cg02560388	13	4	18	0	0.76	1.00	0.89	0.99	13	4	18	0	0.76	1.00	0.89	0.99
TNFSF10	cg01726890	13	4	18	0	0.76	1.00	0.89	0.99	13	4	18	0	0.76	1.00	0.89	0.99
TNFSF10	SLC25A13	13	4	18	0	0.76	1.00	0.89	0.99	13	4	18	0	0.76	1.00	0.89	0.99
SOCS3	LOXL2	15	2	18	0	0.88	1.00	0.94	0.99	15	2	18	0	0.88	1.00	0.94	0.99

Table 6.7: Top performing models from all 65 Bonferroni significant probes. T/F/P/N - true, false, positive, negative. Sens - sensitivity, Spec - specificity, Acc - accuracy, AUC - area under the curve.

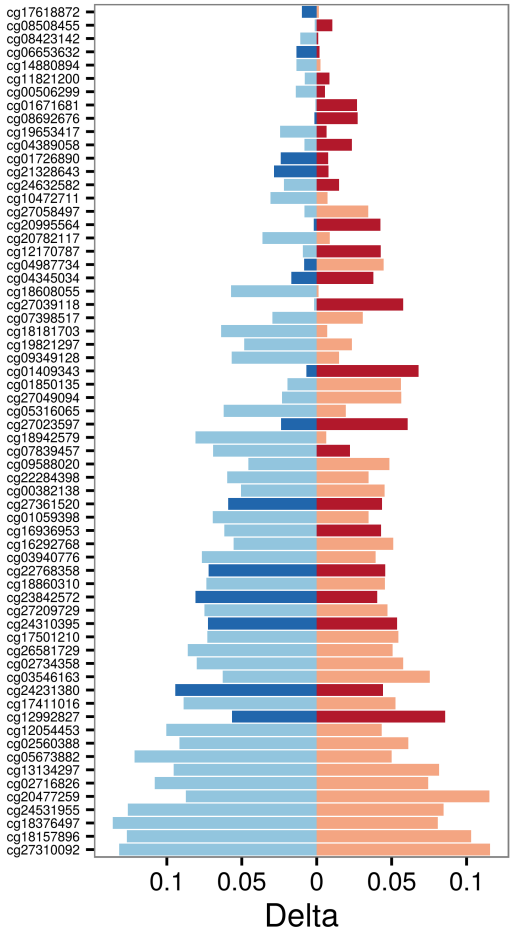


Figure 6.8: Delta - absolute difference in median between adults and children for controls (blue) and Crohn's disease (red). Darker colours indicate relative hypermethylation in adults compared to children, and lighter colours the reverse.

Probes		Methylation								Methylation, age & sex							
		TP	FN	TN	FP	Sens	Spec	Acc	AUC	TP	FN	TN	FP	Sens	Spec	Acc	AUC
RPS6KA2	cg09349128	142	8	94	208	0.95	0.31	0.52	0.81	128	22	172	130	0.85	0.57	0.66	0.82
RPS6KA2	NMRAL1	126	24	188	114	0.84	0.62	0.69	0.81	89	61	263	39	0.59	0.87	0.78	0.81
RPS6KA2	cg12992827	107	43	233	69	0.71	0.77	0.75	0.81	72	78	279	23	0.48	0.92	0.78	0.81
RPS6KA2	TNF	108	42	242	60	0.72	0.80	0.77	0.81	66	84	289	13	0.44	0.96	0.79	0.81
CLU	SBNO2	121	29	199	103	0.81	0.66	0.71	0.81	125	25	179	123	0.83	0.59	0.67	0.81
RPS6KA2	cg04345034	130	20	166	136	0.87	0.55	0.65	0.81	95	55	245	57	0.63	0.81	0.75	0.81
RPS6KA2	PTDSS2	136	14	127	175	0.91	0.42	0.58	0.80	101	49	232	70	0.67	0.77	0.74	0.81
RPS6KA2	ZEB2	128	22	162	140	0.85	0.54	0.64	0.80	104	46	232	70	0.69	0.77	0.74	0.81
RPS6KA2	RFTN1	130	20	138	164	0.87	0.46	0.59	0.80	101	49	238	64	0.67	0.79	0.75	0.81
RPS6KA2	CFI	139	11	103	199	0.93	0.34	0.54	0.80	117	33	188	114	0.78	0.62	0.67	0.81
RPS6KA2	HK2	118	32	210	92	0.79	0.70	0.73	0.79	95	55	256	46	0.63	0.85	0.78	0.81
CLU	SBNO2	132	18	137	165	0.88	0.45	0.60	0.80	124	26	167	135	0.83	0.55	0.64	0.80
RPS6KA2	cg19821297	137	13	123	179	0.91	0.41	0.58	0.80	98	52	240	62	0.65	0.79	0.75	0.80
cg09349128	TNF	112	38	234	68	0.75	0.77	0.77	0.81	127	23	155	147	0.85	0.51	0.62	0.80
cg09349128	MPRIP	140	10	80	222	0.93	0.26	0.49	0.80	142	8	65	237	0.95	0.22	0.46	0.80
RPS6KA2	PTGER4	134	16	133	169	0.89	0.44	0.59	0.80	99	51	235	67	0.66	0.78	0.74	0.80
cg09349128	CLU	131	19	125	177	0.87	0.41	0.57	0.80	132	18	113	189	0.88	0.37	0.54	0.80
RPS6KA2	SLC10A6	141	9	97	205	0.94	0.32	0.53	0.80	107	43	217	85	0.71	0.72	0.72	0.81
RPS6KA2	cg21328643	137	13	125	177	0.91	0.41	0.58	0.80	105	45	223	79	0.70	0.74	0.73	0.80
cg09349128	CDC42BPB	121	29	200	102	0.81	0.66	0.71	0.80	107	43	243	59	0.71	0.80	0.77	0.80
RPS6KA2	ZBTB16	147	3	24	278	0.98	0.08	0.38	0.79	141	9	68	234	0.94	0.23	0.46	0.80
RPS6KA2	cg07398517	134	16	125	177	0.89	0.41	0.57	0.79	98	52	240	62	0.65	0.79	0.75	0.80
RPS6KA2	CLU	139	11	107	195	0.93	0.35	0.54	0.80	94	56	251	51	0.63	0.83	0.76	0.80
RPS6KA2	MPRIP	143	7	71	231	0.95	0.24	0.47	0.79	107	43	209	93	0.71	0.69	0.70	0.80
cg12992827	CFI	97	53	247	55	0.65	0.82	0.76	0.80	86	64	260	42	0.57	0.86	0.77	0.80
RPS6KA2	PRKCE	143	7	88	214	0.95	0.29	0.51	0.80	112	38	205	97	0.75	0.68	0.70	0.80
RPS6KA2	ARHGEF3	130	20	162	140	0.87	0.54	0.65	0.78	99	51	251	51	0.66	0.83	0.77	0.81
cg09349128	NMRAL1	117	33	208	94	0.78	0.69	0.72	0.80	133	17	144	158	0.89	0.48	0.61	0.79
RPS6KA2	RUNX3	131	19	155	147	0.87	0.51	0.63	0.79	84	66	267	35	0.56	0.88	0.78	0.80

Table 6.8: Top performing models built from the combined paediatric dataset, tested in the IBD Character adult cohort. T/F/P/N - true, false, positive, negative. Sens - sensitivity, Spec - specificity, Acc - accuracy, AUC - area under the curve.

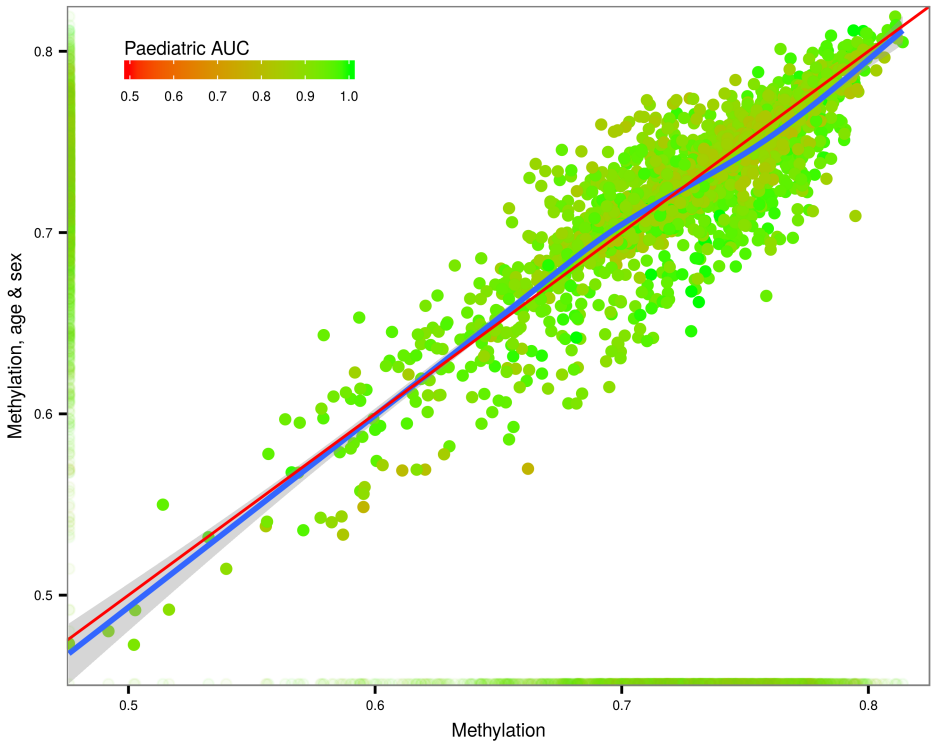


Figure 6.9: Area under the curve for adult replication of paediatric biomarkers with and without age and sex as covariates, based on the complete paediatric dataset. Point colour shows the AUC of the biomarker in the paediatric replication cohort when constructed with the paediatric discovery cohort. Density indicated on each axis.

6.4 Discussion

Paediatric results No single methylation result could distinguish between CD and symptomatic controls with enough accuracy for clinical use. However, a combination of two methylation results was highly accurate — up to 94% using only the probes which were significant in the initial analysis of the paediatric replication cohort, or up to 100% using the probes which were significant from the combined analysis, though this is less rigorous.

These findings were not restricted to particular probes and although some models performed better than others, any possible combination of two significant probes proved effective, with median results as shown in table 6.4 (page 239).

Differences in the absolute methylation values were observed between adults and children, in spite of similar relative methylation changes associated with disease (figure 6.8, page 247) which resulted in a substantial reduction in the accuracy of the models . This may be due to differences in the proportions of cellular populations between adults and children, or an independent effect of ageing.

The quoted significance of results are adjusted for cell type. This entails measuring or predicting the proportions of various cell types within each sample, and including these proportions as covariates when estimating the association between disease and methylation.

When correcting for cell type in the paediatric dataset, we found that the publicly available reference dataset of Illumina 450k from separated cells functioned poorly to predict whole blood makeup in children[3]. We therefore used the full blood count data we had for the majority of patients, to find probes which did not vary by disease but accurately predicted the differential cell counts in these patients, and then applied this model to predict the cell counts in the minority of patients in whom we lacked data. This strengthens the indication that there are fundamental age-related differences, which would make these biomarkers of less use in adults. However, the individual probes which exhibited methylation changes, the direction, and magnitude of those changes do correlate well between children and adults, and so for a future study which aimed to develop biomarkers applicable to children and adults, a learning dataset covering both children and adults would be needed.

[3]work by NAK

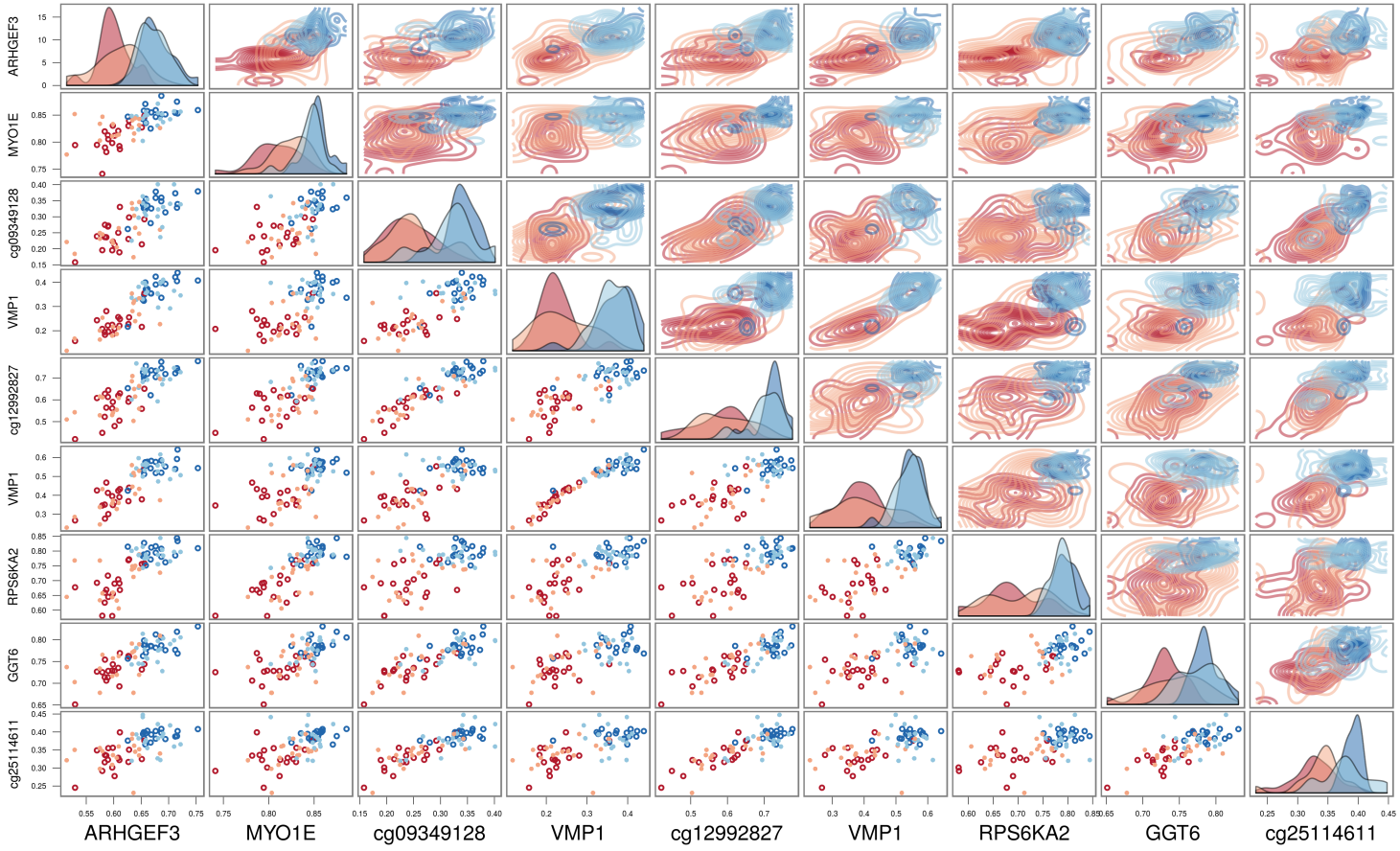


Figure 6.10: Results in all paediatric samples for top probes in the paediatric discovery cohort used to train models; lower-left - individual β values for each two-probe combination; upper-right - 2D probability density; diagonal - 1D probability density. Red: CD, blue: control, darker: discovery, lighter: replication

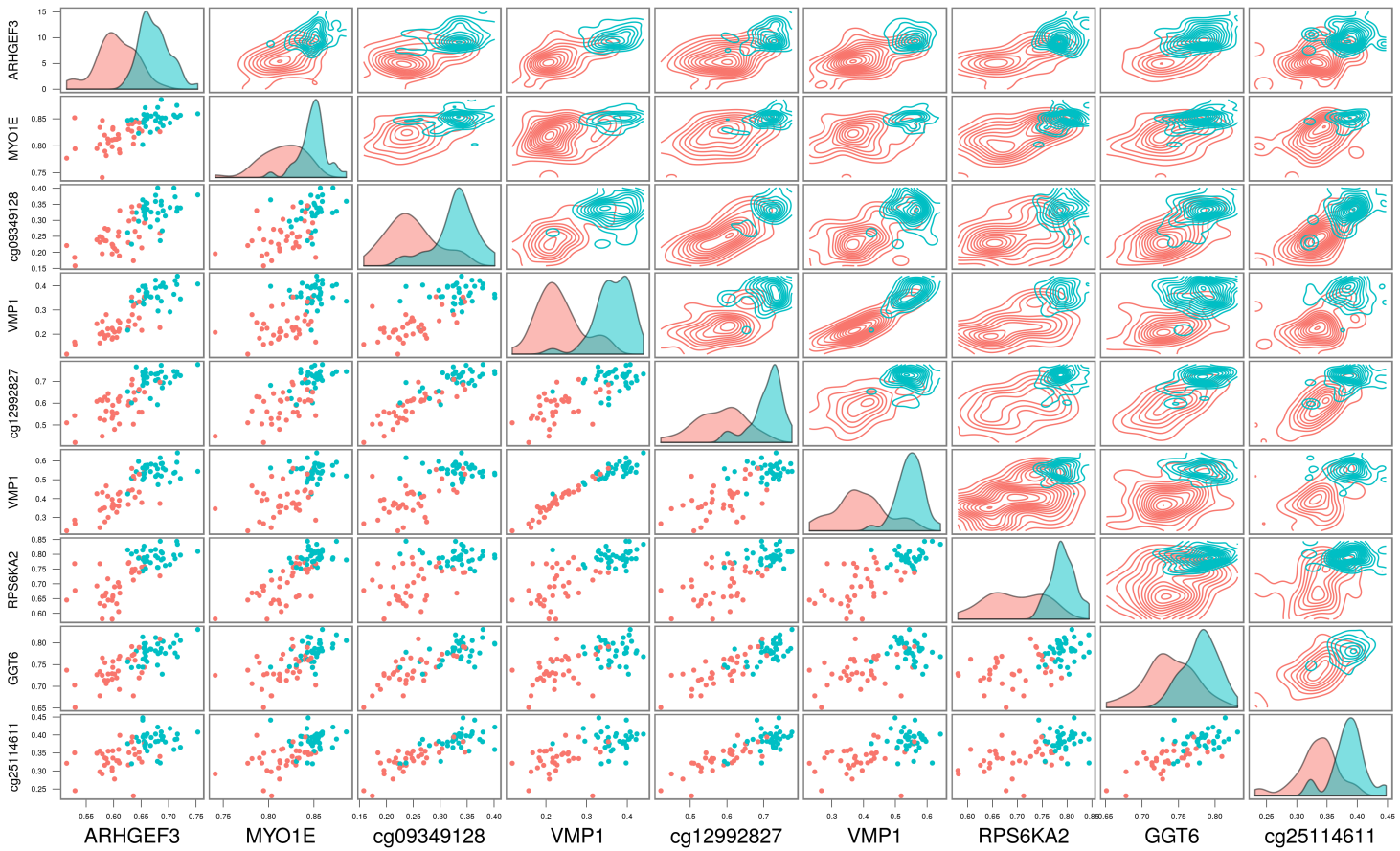


Figure 6.11: Results in all paediatric samples for top probes in the paediatric discovery cohort used to train models; lower-left - individual β values for each two-probe combination; upper-right - 2D probability density by diagnosis; diagonal - 1D probability density. Red: CD, blue: control.

Appendix A

IBD Character

A.1 Introduction

The Inflammatory Bowel Disease characterisation by a multi-modal integrated biomarker study consortium ("IBD Character") is an ongoing multi-centre study funded by the European Commission's seventh framework programme for research and technological development (FP7). The consortium will be performing multi-omics analysis, including genotyping, RNA and protein expression, DNA methylation and microbiome analysis in a variety of tissues and sample types. The aim is to analyse these data to develop new biomarkers for the diagnosis and management of IBD, and to identify novel therapeutic targets. The patients recruited into IBD Character are recently diagnosed and treatment naive.

The IBD Character project has not yet finished, with final phenotype data expected in April 2016. Some preliminary results from the Illumina 450k analysis of whole blood methylation have been reported in this thesis as a replication of the paediatric results. An abstract detailing preliminary methylation findings from analysis of the IBD Character cohort has been accepted for presentation at the United European Gastroenterology Week conference in October 2016 (Appendix B.3, page 293) and publication thereafter.

The centres contributing samples and sample numbers are shown in table A.1.

This appendix contains preliminary data from Illumina 450k analysis of the IBD-Character project, and demonstrates several aspects of Illumina 450k QC and analysis, provides an independent replication of the paediatric Illumina 450k results presented in chapter 3, and shows that similar results are obtained from IBD patients and controls from outwith the UK.

A.2 Materials & methods

Methylation was assayed using the Illumina 450k platform. Samples containing >1% probes with detection p values >0.05 were removed. Probes with bead counts <3 in 5% of samples, or detection p values >0.05 in 1% of samples were also removed.

Normalisation was performed using the 'dasen' method as implemented in the 'wateRmelon' R package, which involves equalisation of type I and II probe backgrounds, followed by inter-array quantile normalisation of type I and type II probes separately [493].

Sex mismatches were identified by analysis of median intensities of sex chromosome probes, and removed from further analysis.

Proportions for CD8 T cells, CD4 T cells, B-cells, monocytes, granulocytes and NK cells were estimated using the R package Minfi [25] based on the work of Houseman *et al.* [270].

Probes containing SNPs were identified by the manufacturer[1], but annotated rather than filtered.

Disease associated methylation changes were identified by constructing linear models with the R package 'limma' [582] using cell proportions as covariates.

Differentially methylated regions were defined using the 'lasso' function in the R package 'ChAMP' [443].

Genotyping was performed using the Illumina HumanOmniExpressExome-8 platform. Methylation eQTLs were found using the 'MatrixEQTL' R package [563] using a distance threshold of 1Mb for cis-meqtls.

[1] https://support.illumina.com

Centre	CD	Control	IBDU	UC	Unknown	Total
Edinburgh	47	87	11	58	3	206
Linkoping	9	39	0	3	0	51
Maastricht	4	0	0	1	0	5
Orebro	24	64	8	25	5	126
Oslo	52	68	5	70	1	196
Zaragoza	14	40	2	10	0	66
Total	150	298	26	167	9	650

Table A.1: IBD Character samples by centre and diagnosis, after removal of samples failing QC

A.3 Results

A.3.1 QC

Three samples (0.4% of total) with low signal intensities were removed having >1% of probes with detection p values >0.05, and 3040 probes (0.6%) were removed due to low signal.

Five clear sex mismatches (0.7%) were identified, as shown in figure A.1, and removed. Two additional male samples with low Y chromosome signal intensity are also evident (sample IDs 2113 and 4114).

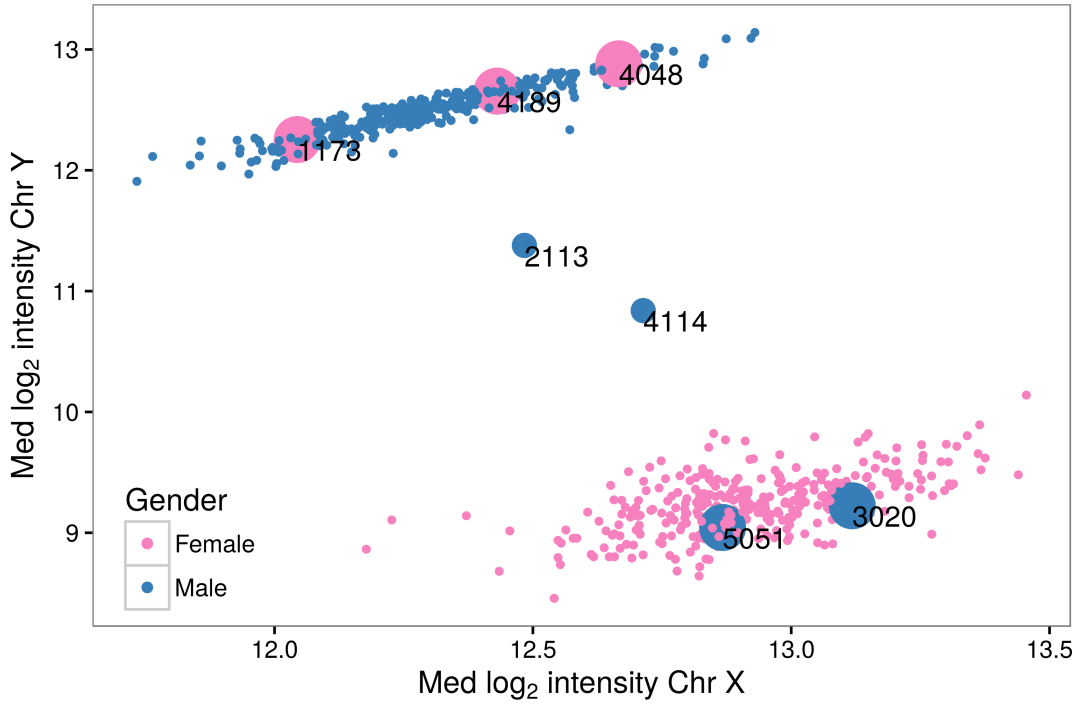


Figure A.1: Sex mismatch identification by median X and Y chromosome signal intensity

A.3.2 Replication between Edinburgh and non-Edinburgh samples

Edinburgh was the largest individual centre, with 32% of all samples. As the character results have been used to replicate findings in chapters 4–6 it was important to ensure that these results were not due to the Edinburgh cohort alone. Results were therefore compared between Edinburgh and and other centres, which revealed a very strong correlation as shown in figure A.2 (Spearman's ρ=0.94, p<2.2×10^{-308}, top 1000 probes).

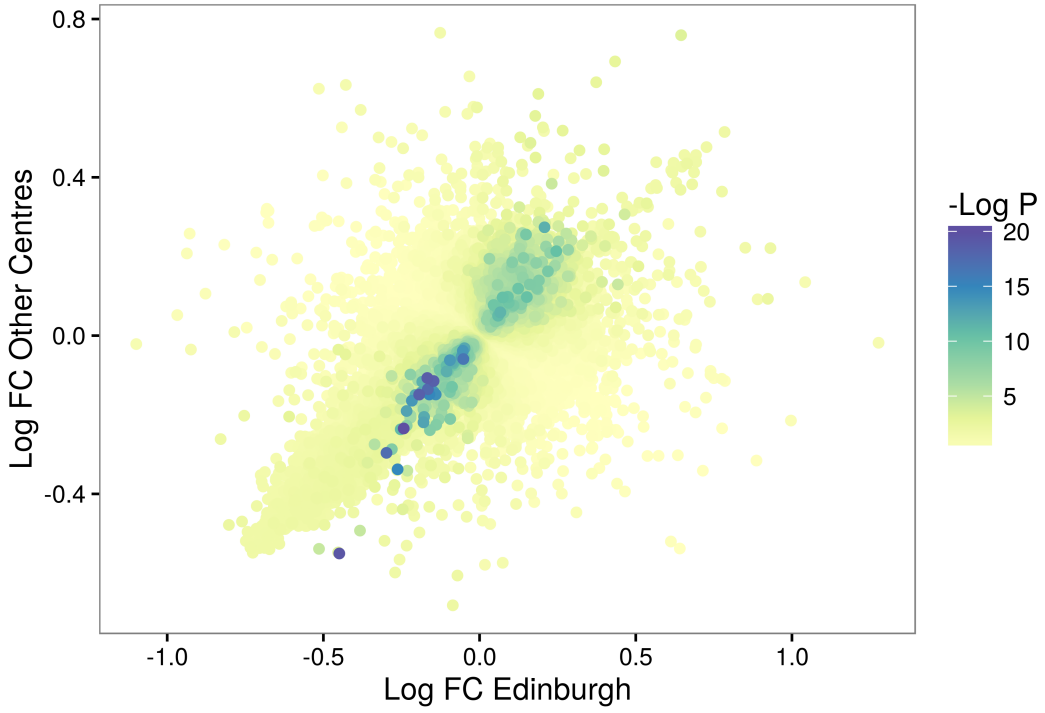


Figure A.2: Log fold-change in Edinburgh vs other centres

A.3.3 Individual CpGs

The number of Bonferroni significant findings for each analysis are shown in table A.2, followed by the top 20 results for each on pages 259–260. Probes containing SNPs are indicated, and MAF shown. The complete results are shown in section A.4 (from page 271).

Comparison	Significant probes
IBD vs control	193
CD vs control	81
UC vs control	261
CD vs UC	1

Table A.2: Number of Bonferroni significant results for each performed analysis

Probe	Chr	Coord	Gene	P value	SNP MAF
cg09349128	22	50327986	CRELD2	8.4×10^{-21}	0.027
cg12054453	17	57915717	VMP1	3.7×10^{-20}	
cg17501210	6	166970252	RPS6KA2	1.1×10^{-19}	
cg12170787	19	1130965	SBNO2	2.7×10^{-19}	
cg18608055	19	1130866	SBNO2	2.8×10^{-19}	
cg07573872	19	1126342	SBNO2	2.2×10^{-18}	
cg16936953	17	57915665	VMP1	4.1×10^{-18}	
cg27469606	19	1154485	SBNO2	2.7×10^{-17}	
cg01059398	3	172235808	TNFSF10	1.1×10^{-15}	
cg02650017	17	47301614	PHOSPHO1	1.3×10^{-15}	
cg18513344	3	195531298	MUC4	5.5×10^{-15}	
cg03156547	14	23525003	CDH24	1.7×10^{-14}	
cg24002003	15	101668143	CHSY1	1.8×10^{-14}	
cg02716826	9	33447032	AQP3	1.9×10^{-14}	
cg06164260	3	187454439	BCL6	2.9×10^{-14}	
cg13781414	9	138951648	NACC2	8.5×10^{-14}	
cg10636246	1	159046973	AIM2	1.7×10^{-13}	
cg18942579	17	57915773	VMP1	1.8×10^{-13}	
cg01983725	1	227161799	ADCK3	4.2×10^{-13}	
cg19821297	19	12890029	HOOK2	5.7×10^{-13}	

Table A.3: IBD Character: IBD vs control - top 20

Probe	Chr	Coord	Gene	P value	SNP MAF
cg12170787	19	1130965	SBNO2	1.4×10^{-22}	
cg18608055	19	1130866	SBNO2	5.5×10^{-21}	
cg09349128	22	50327986	CRELD2	1.1×10^{-18}	0.027
cg12054453	17	57915717	VMP1	1.6×10^{-17}	
cg17501210	6	166970252	RPS6KA2	4.3×10^{-17}	
cg07573872	19	1126342	SBNO2	9.5×10^{-17}	
cg16936953	17	57915665	VMP1	1.7×10^{-14}	
cg27469606	19	1154485	SBNO2	2.2×10^{-14}	
cg01059398	3	172235808	TNFSF10	3.1×10^{-12}	
cg10967866	10	134362164	INPP5A	3.2×10^{-12}	
cg18513344	3	195531298	MUC4	4.3×10^{-12}	
cg15551881	9	123688715	TRAF1	8.3×10^{-12}	
cg20141108	1	165907859	UCK2	9.6×10^{-12}	
cg19821297	19	12890029	HOOK2	1.0×10^{-11}	
cg13781414	9	138951648	NACC2	1.2×10^{-11}	
cg18181703	17	76354621	SOCS3	1.3×10^{-11}	
cg02650017	17	47301614	PHOSPHO1	2.1×10^{-11}	
cg16292768	8	27467783	CLU	3.5×10^{-11}	
cg18942579	17	57915773	VMP1	6.7×10^{-11}	
cg03940776	6	158490013	SYNJ2	1.1×10^{-10}	

Table A.4: IBD Character: CD vs control - top 20

Probe	Chr	Coord	Gene	P value	SNP MAF
cg09349128	22	50327986	CRELD2	1.3×10^{-17}	0.027
cg16936953	17	57915665	VMP1	1.5×10^{-15}	
cg12054453	17	57915717	VMP1	4.0×10^{-15}	
cg17501210	6	166970252	RPS6KA2	6.7×10^{-15}	
cg07573872	19	1126342	SBNO2	8.5×10^{-15}	
cg24002003	15	101668143	CHSY1	1.2×10^{-14}	
cg18608055	19	1130866	SBNO2	7.0×10^{-14}	
cg08449144	11	1857919	SYT8	9.2×10^{-14}	
cg06164260	3	187454439	BCL6	1.1×10^{-13}	
cg02650017	17	47301614	PHOSPHO1	1.6×10^{-13}	
cg27469606	19	1154485	SBNO2	2.0×10^{-13}	
cg01059398	3	172235808	TNFSF10	2.7×10^{-13}	
cg02716826	9	33447032	AQP3	7.3×10^{-13}	
cg10636246	1	159046973	AIM2	1.6×10^{-12}	
cg01983725	1	227161799	ADCK3	1.8×10^{-12}	
cg03156547	14	23525003	CDH24	2.4×10^{-12}	
cg20995564	2	145172035	ZEB2	3.5×10^{-12}	
cg03039243	1	2782699	TTC34	5.9×10^{-12}	
cg02734358	4	90227074	GPRIN3	6.3×10^{-12}	
cg12170787	19	1130965	SBNO2	6.7×10^{-12}	

Table A.5: IBD Character: UC vs control - top 20

A.3.4 Comparison with paediatric results

In the paediatric analysis there were 65 probes with significant disease-associated methylation differences, one of these probes was not analysed in the IBD Character cohort as it failed QC. Of the remaining 64 probes which were significant in children, 22 also achieved Bonferroni significance in adults representing a modest proportion (34.4%) but a massive enrichment compared to chance (2376 times enrichment, $p=5.38\times10^{-7586}$). For the very strongest results the concordance was higher; the top five paediatric results were all significant in the adults, and the top 6 adult results were all significant in the children. Figure A.3 shows the Log_2 fold-change in the paediatric results and IBD Character results.

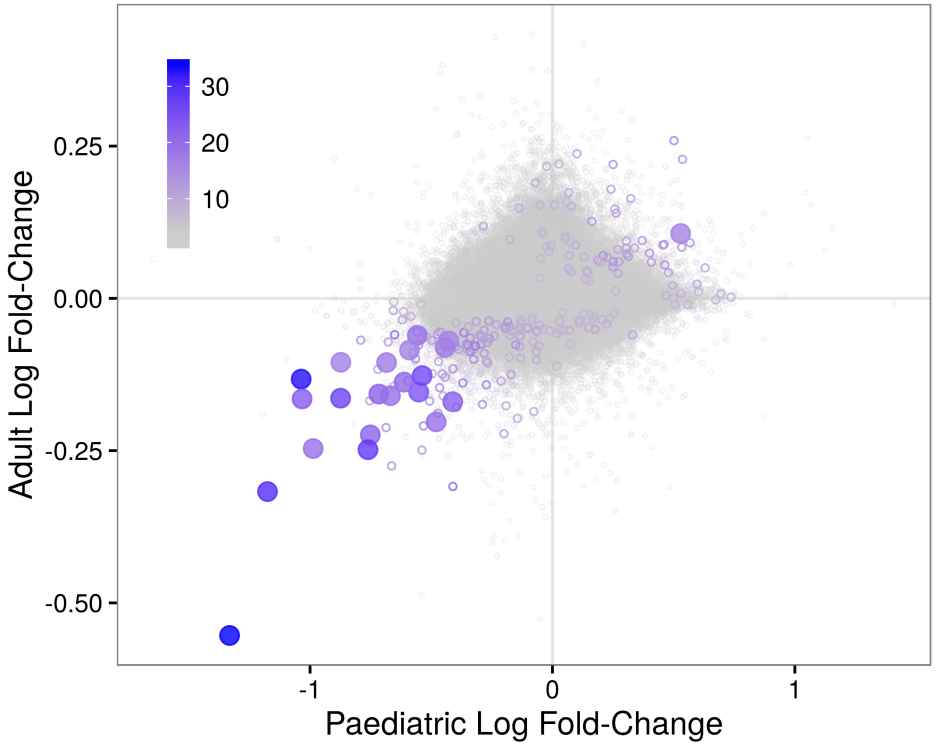


Figure A.3: Log_2 fold-change in paediatric and IBD Character samples. Small points were significant in neither cohort, medium open circles in one cohort, and large filled circles in both cohorts. Coloured by the sum of the $-Log_{10}$ p values

A.3.5 Differentially methylated regions

Using the 'Lasso' function in the R package 'ChAMP' [443] 1709 differentially methylated regions of consecutive FDR significant probes were defined. Figure A.4 shows the variability in probe density in different genomic environments as found in the IBD Character dataset, and figures A.5 and A.6 (pages 263–264) show the resulting lasso radii. Table A.6 (267) show the 100 most significant DMRs as calculated by Stouffler's method of p value aggregation. These include *VMP1*, *RUNX3*, *ITGB2* and *DIABLO*, replicating findings in the paediatric dataset, and numerous results throughout the HLA region.

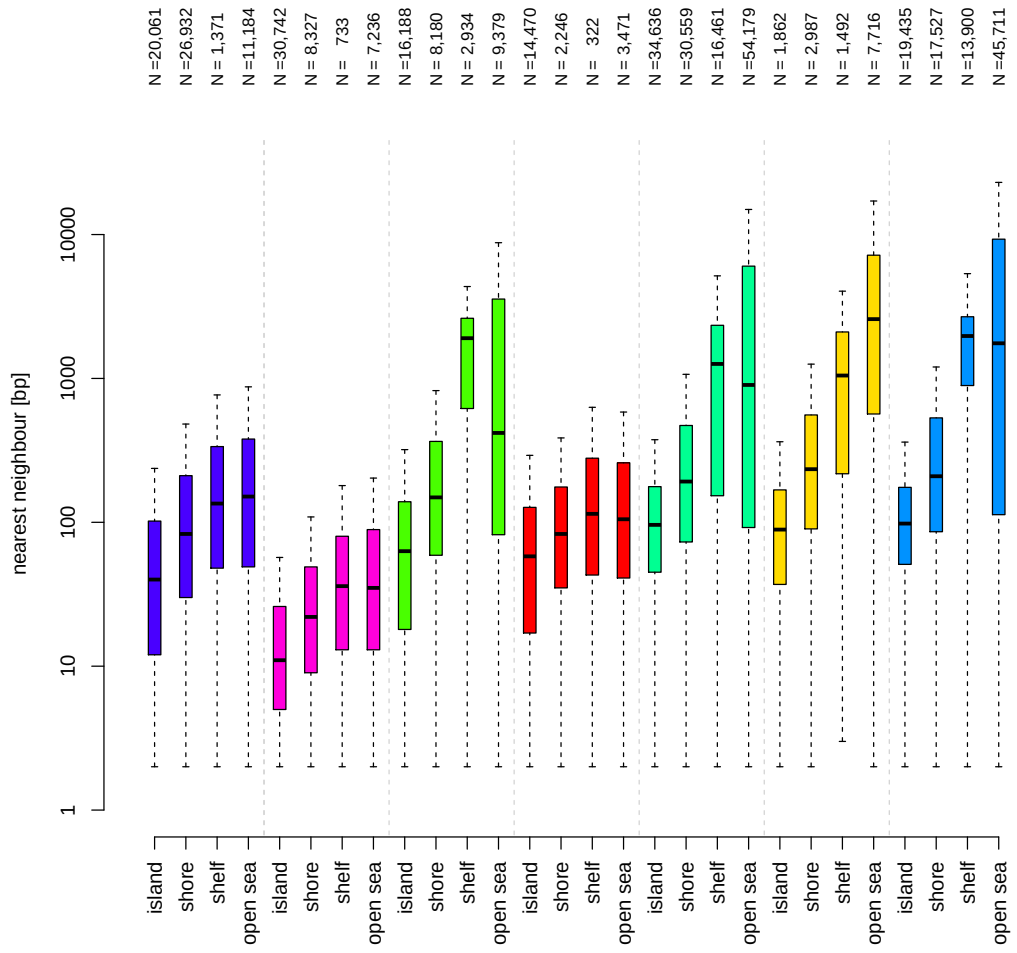


Figure A.4: Distance to the nearest probe by location (from left to right: TSS1500, TSS200, 5′UTR, 1st exon, gene body, 3′UTR, and intergenic region. Colour key in figure A.4), and by CpG density (island, shore, shelf, or open sea).

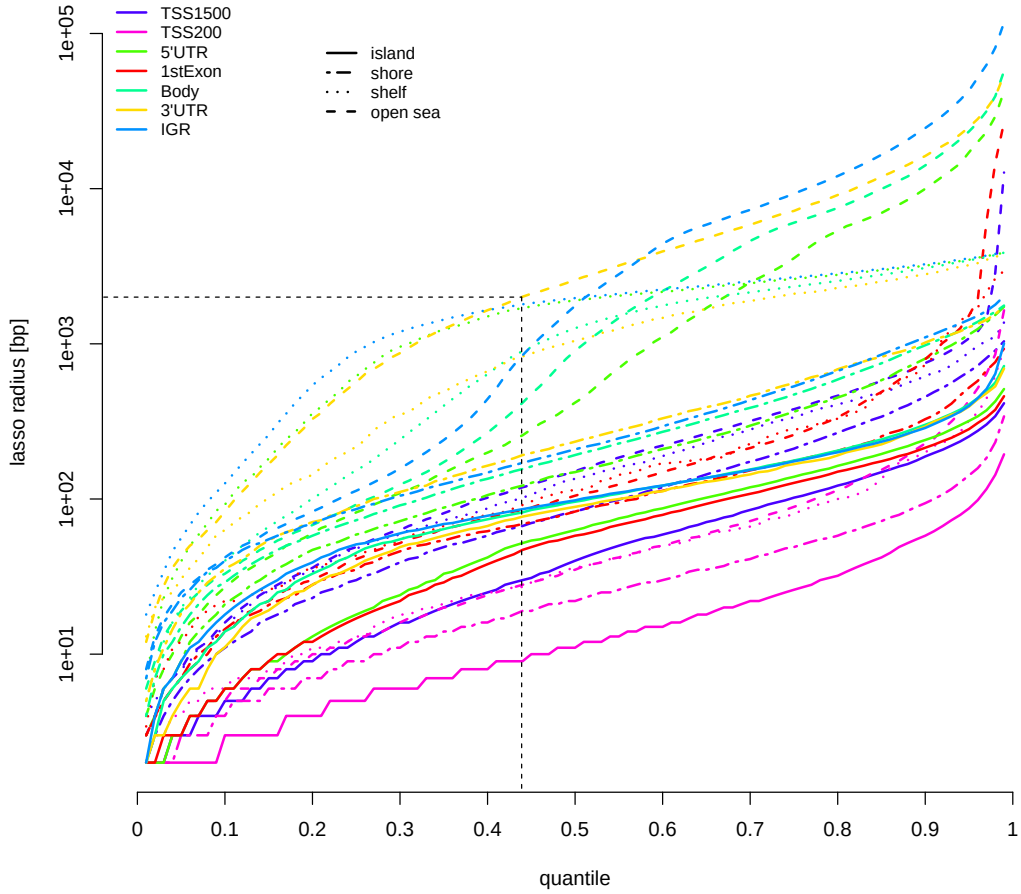


Figure A.5: The lasso radii needed for each region type to include a given proportion of probes. The chosen quantile of 0.44 is indicated.

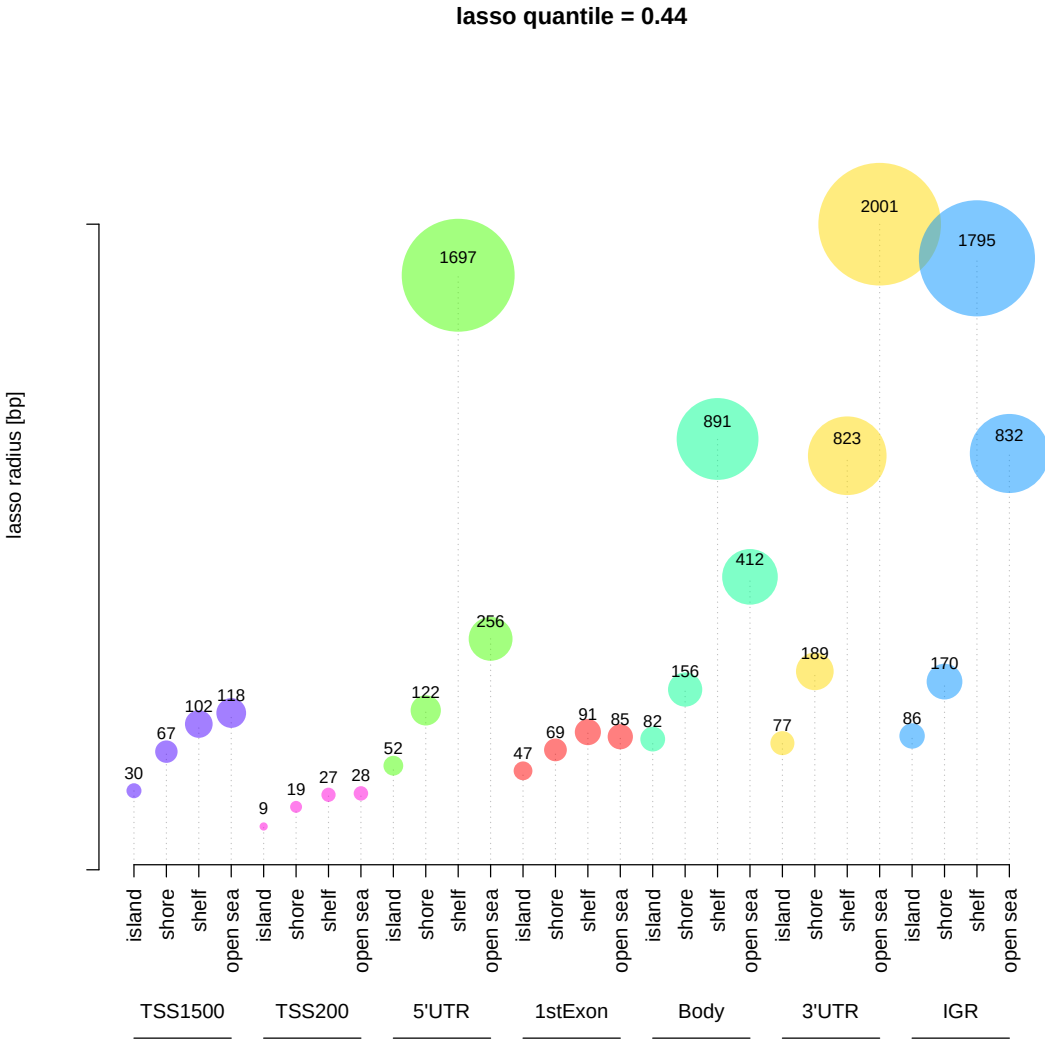


Figure A.6: The range of lasso radii for each region type and CpG density at the chosen quantile of 0.44

Chr	Coordinate	Gene	Size	β_{CD}-β_{HC}	P value
6	31539945	LTA	1218	0.040	1.19×10^{-198}
5	148808075	MIR143	2164	-0.036	4.66×10^{-123}
1	25291387	RUNX3	804	0.034	1.22×10^{-122}
10	49892711	WDFY4	1095	-0.031	1.74×10^{-116}
14	22974119	DAD1	5443	0.027	1.59×10^{-99}
3	52528946	STAB1	991	-0.036	1.91×10^{-94}
11	118213074	CD3D	1890	0.037	6.88×10^{-93}
12	122711832	DIABLO	313	0.039	2.87×10^{-91}
20	57582789	CTSZ	288	-0.054	3.95×10^{-90}
13	40761318	LINC00332	1984	0.036	5.23×10^{-85}
1	32715264	LCK	4054	0.026	5.68×10^{-78}
21	43823691	UBASH3A	984	0.040	1.50×10^{-76}
2	33358786	LTBP1	983	-0.054	5.13×10^{-72}
1	11795830	AGTRAP	175	-0.048	1.02×10^{-71}
18	21452612	LAMA3	237	0.035	3.78×10^{-69}
20	45179072	C20orf123	460	-0.053	1.45×10^{-67}
10	63808661	ARID5B	922	0.029	6.95×10^{-67}
16	66400202	CDH5	624	-0.045	3.02×10^{-66}
13	34184100	RFC3	2044	0.033	1.93×10^{-65}
11	843830	TSPAN4	3923	-0.028	6.43×10^{-65}
17	62083949	ICAM2	667	-0.035	2.07×10^{-64}
12	111618513	CUX2	1355	-0.024	5.02×10^{-64}
7	81399052	HGF	747	-0.037	7.02×10^{-64}
22	44568219	PARVG	585	-0.040	8.05×10^{-64}
22	45072199	PRR5	1223	-0.030	1.04×10^{-63}
1	154941603	SHC1	4027	-0.034	1.60×10^{-62}
18	21572492	TTC39C	513	0.032	3.34×10^{-62}
11	69259077	CCND1	6413	-0.033	6.57×10^{-62}
11	94278366	FUT4	285	-0.032	3.56×10^{-61}
13	114829084	RASA3	2042	-0.028	1.37×10^{-59}
9	35649015	SIT1	2438	0.034	8.58×10^{-59}
1	153330143	S100A9	1046	-0.040	8.74×10^{-59}
16	3115015	IL32	949	0.026	8.74×10^{-59}
19	13052828	CALR	2498	-0.045	8.74×10^{-59}
7	105319025	ATXN7L1	1207	-0.041	2.38×10^{-58}
4	48135978	TXK	513	0.047	2.97×10^{-57}
1	203734147	LAX1	825	0.039	4.14×10^{-57}

Table A.6: DMRs: Continued overleaf

Chr	Coordinate	Gene	Size	β_{CD}-β_{HC}	P value
17	79128027	AATK	1943	-0.071	2.40×10^{-56}
17	33775914	SLFN13	69	-0.043	4.38×10^{-56}
7	98738107	SMURF1	1832	0.025	9.87×10^{-55}
17	57915253	VMP1	4338	-0.011	1.12×10^{-54}
12	56414324	IKZF4	237	-0.051	1.51×10^{-54}
12	2733738	CACNA1C	1266	-0.040	2.93×10^{-54}
7	128579095	IRF5	3395	-0.037	3.15×10^{-54}
8	141598729	EIF2C2	1120	0.033	1.01×10^{-53}
15	29213508	APBA2	609	0.023	2.44×10^{-53}
22	42827713	NFAM1	825	-0.024	2.44×10^{-53}
7	5564516	ACTB	3591	0.024	4.28×10^{-53}
17	76129246	TMC8	2142	0.024	1.03×10^{-52}
6	30719248	IER3	2076	-0.041	1.71×10^{-51}
6	31543455	TNF	3454	0.004	1.98×10^{-51}
8	144543156	ZC3H3	1208	0.038	1.98×10^{-51}
8	145024977	PLEC1	284	-0.039	3.20×10^{-51}
17	75314825	SEPT9	1912	-0.026	3.23×10^{-51}
10	72362667	PRF1	245	0.032	1.62×10^{-50}
8	28917989	KIF13B	2078	0.038	1.64×10^{-50}
11	118754309	CXCR5	541	0.021	3.26×10^{-50}
11	67204916	PTPRCAP	2410	0.025	3.72×10^{-50}
17	19289941	MFAP4	940	-0.028	4.36×10^{-50}
12	122443618	BCL7A	1836	0.042	4.74×10^{-50}
1	9788526	PIK3CD	433	0.049	1.16×10^{-49}
2	242801997	PDCD1	205	0.046	2.13×10^{-49}
3	128779558	GP9	72	-0.058	8.52×10^{-49}
22	38714253	CSNK1E	281	-0.030	1.22×10^{-48}
21	46346609	ITGB2	3780	0.025	1.84×10^{-48}
1	167486883	CD247	1163	0.050	2.04×10^{-48}
1	24861795	RCAN3	155	0.048	7.11×10^{-48}
11	818834	PNPLA2	153	-0.048	1.55×10^{-47}
17	78753344	RPTOR	2673	-0.032	2.27×10^{-47}
11	57528999	CTNND1	872	-0.047	2.55×10^{-47}
19	827758	AZU1	825	-0.040	5.01×10^{-47}
17	9939592	GAS7	942	-0.066	8.05×10^{-47}
17	2907266	RAP1GAP2	1516	-0.027	1.37×10^{-46}
17	46622473	HOXB2	4655	-0.029	1.81×10^{-46}

Table A.6: DMRs: Continued overleaf

Chr	Coordinate	Gene	Size	β_{CD}-β_{HC}	P value
11	36422121	PRR5L	751	-0.030	3.78×10^{-46}
7	23386953	IGF2BP3	856	-0.061	1.11×10^{-45}
14	101908582	DIO3OS	341	0.033	2.78×10^{-45}
12	53612485	RARG	313	0.024	2.82×10^{-45}
7	139319934	HIPK2	986	0.036	3.76×10^{-45}
19	847924	PRTN3	225	-0.042	5.57×10^{-45}
7	2772854	GNA12	1180	-0.040	6.42×10^{-45}
2	68592260	PLEK	890	-0.033	9.39×10^{-45}
17	27044957	RAB34	376	-0.038	1.10×10^{-44}
2	241642371	KIF1A	3665	0.038	1.11×10^{-44}
5	963398	LOC100506688	2089	0.037	1.95×10^{-44}
12	124864116	NCOR2	978	-0.047	3.13×10^{-44}
11	1951787	TNNT3	1861	-0.032	3.37×10^{-44}
11	64635316	EHD1	943	0.033	6.80×10^{-44}
16	85981140	IRF8	404	0.038	1.02×10^{-43}
17	56355217	MPO	228	-0.051	1.77×10^{-43}
6	158066197	ZDHHC14	1116	-0.054	2.01×10^{-43}
1	211501522	TRAF5	3395	0.026	2.19×10^{-43}
16	4713668	MGRN1	1560	-0.031	6.30×10^{-43}
21	34771577	IFNGR2	3591	-0.039	6.80×10^{-43}
5	154229885	C5orf4	545	-0.049	1.49×10^{-42}
3	15310609	SH3BP5	1073	-0.034	1.67×10^{-42}
1	6526464	TNFRSF25	135	0.033	2.86×10^{-42}
10	97515254	ENTPD1	263	-0.066	3.49×10^{-42}
17	3704386	ITGAE	245	-0.040	6.00×10^{-42}
11	113933403	ZBTB16	1937	0.018	6.01×10^{-42}

Table A.6: Top 100 DMRs between IBD and control in the IBD Character cohort. Coordinates as per Hg19, coordinate refers to the smallest coordinate of the DMR bounds, size denotes the width of the DMR in bp. β_{CD}-β_{HC} shows the difference in mean βvalues by diagnosis. P values as calculated by the ChAMP package using Stouffler's method of p value aggregation.

A.3.6 Methylation quantitative trait loci

After FDR correction, there were 2327 cis-meQTLs. The 100 most significant of these are shown in table A.7.

SNP	Probe	Gene	P value
rs1932819	cg03990033	CD34	2.06×10^{-108}
rs1109278	cg10460946	LOC100996291	1.47×10^{-97}
rs1109279	cg10460946	LOC100996291	3.37×10^{-89}
rs7567627	cg18991601	FZD5	5.93×10^{-84}
rs1567237	cg10460946	LOC100996291	2.21×10^{-79}
rs886926	cg12582317	MSI2	5.29×10^{-72}
rs9893344	cg12582317	MSI2	2.70×10^{-71}
rs6755821	cg18991601	FZD5	3.54×10^{-67}
rs9657682	cg21240736	LINC00963	5.22×10^{-66}
rs2663879	cg18991601	FZD5	6.46×10^{-65}
rs2663864	cg18991601	FZD5	1.04×10^{-62}
rs3754171	cg03990033	CD34	5.29×10^{-59}
rs2267894	cg03990033	CD34	1.21×10^{-58}
rs11666543	cg08363114	RPS9	1.54×10^{-50}
rs907605	cg08449144	SYT8	9.01×10^{-50}
rs2663887	cg18991601	FZD5	4.61×10^{-46}
rs9500284	cg07967630	GUSBP4	6.01×10^{-46}
rs1891818	cg07967630	GUSBP4	7.52×10^{-43}
rs2070946	cg18663307	ITGB2	3.14×10^{-42}
rs2070946	cg04321224	ITGB2	5.69×10^{-42}
rs3741231	cg08449144	SYT8	7.00×10^{-40}
rs3900115	cg12105450	CASP10	8.01×10^{-39}
rs9892194	cg12582317	MSI2	1.20×10^{-38}
rs415890	cg24160243	RPS6KA2	1.03×10^{-37}
rs12693932	cg12105450	CASP10	2.25×10^{-37}
rs394522	cg24160243	RPS6KA2	3.74×10^{-37}
rs3834129	cg12105450	CASP10	1.16×10^{-35}
rs12693931	cg12105450	CASP10	1.39×10^{-35}
rs1147332	cg21240736	LINC00963	1.58×10^{-35}
rs7576306	cg12105450	CASP10	3.84×10^{-35}
rs10188598	cg12105450	CASP10	1.16×10^{-34}
rs4793573	cg12582317	MSI2	1.37×10^{-34}
rs3769827	cg12105450	CASP10	1.26×10^{-32}

Table A.7: MeQTLs: Continued overleaf

SNP	Probe	Gene	P value
rs1594	cg12105450	CASP10	2.30×10^{-32}
rs12029239	cg03990033	CD34	8.22×10^{-32}
rs2298803	cg17515347	AIM2	6.79×10^{-31}
rs2236313	cg24160243	RPS6KA2	4.64×10^{-29}
rs2236313	cg24160243	RPS6KA2	4.64×10^{-29}
rs9459836	cg24160243	RPS6KA2	2.98×10^{-27}
rs9306118	cg04321224	ITGB2	5.59×10^{-27}
rs11672654	cg08363114	RPS9	6.20×10^{-27}
rs144494015	cg16932065	LIMK2	1.74×10^{-26}
rs9306118	cg18663307	ITGB2	1.92×10^{-26}
rs4506048	cg07967630	GUSBP4	2.40×10^{-26}
rs7818871	cg14722693	CSGALNACT1	3.99×10^{-26}
rs1014482	cg14722693	CSGALNACT1	1.70×10^{-25}
rs10107533	cg14722693	CSGALNACT1	4.22×10^{-25}
rs8078424	cg12054453	VMP1	4.44×10^{-25}
rs2849030	cg12616487	EML3	4.96×10^{-25}
rs11231155	cg12616487	EML3	5.25×10^{-25}
rs2301436	cg24160243	RPS6KA2	5.92×10^{-25}
rs2301436	cg24160243	RPS6KA2	1.25×10^{-24}
rs11231156	cg12616487	EML3	1.57×10^{-24}
rs17009223	cg03990033	CD34	1.70×10^{-24}
rs11231154	cg12616487	EML3	2.43×10^{-24}
rs2273843	cg03875678	GZMB	5.39×10^{-24}
rs2509982	cg12616487	EML3	7.76×10^{-24}
rs3903046	cg07967630	GUSBP4	7.91×10^{-24}
rs2256964	cg15082870	SEPT7	8.10×10^{-24}
rs7749278	cg24160243	RPS6KA2	9.26×10^{-24}
rs7124057	cg12616487	EML3	1.02×10^{-23}
rs11142402	cg00049440	KLF9	1.58×10^{-23}
rs2302479	cg07698196	TMEM176B	2.43×10^{-23}
rs962927	cg15429134	BORCS5	2.70×10^{-23}
rs9645752	cg15429134	BORCS5	5.39×10^{-23}
rs1056448	cg00049440	KLF9	5.70×10^{-23}
rs8078424	cg16936953	VMP1	8.91×10^{-23}
rs10868824	cg00049440	KLF9	1.08×10^{-22}
rs11553576	cg12616487	EML3	1.22×10^{-22}
rs3769825	cg12105450	CASP10	2.81×10^{-22}

Table A.7: MeQTLs: Continued overleaf

SNP	Probe	Gene	P value
rs9650314	cg02508743	LYN	7.44×10^{-22}
rs10845552	cg15429134	BORCS5	1.27×10^{-21}
rs4921659	cg14722693	CSGALNACT1	1.38×10^{-21}
rs2070946	cg24707889	ITGB2	1.52×10^{-21}
rs933575	cg15429134	BORCS5	2.07×10^{-21}
rs2838738	cg18663307	ITGB2	3.20×10^{-21}
rs79885699	cg10460946	LOC100996291	3.39×10^{-21}
rs10051822	cg25368647	MXD3	5.93×10^{-21}
rs7818368	cg14722693	CSGALNACT1	6.50×10^{-21}
rs1147331	cg21240736	LINC00963	6.86×10^{-21}
rs10853015	cg12054453	VMP1	7.45×10^{-21}
rs2667978	cg02508743	LYN	7.49×10^{-21}
rs3018561	cg12616487	EML3	1.06×10^{-20}
rs10281675	cg07698196	TMEM176B	1.76×10^{-20}
rs2256964	cg16805291	SEPT7	1.84×10^{-20}
rs2516633	cg12616487	EML3	2.85×10^{-20}
rs2333562	cg12054453	VMP1	2.97×10^{-20}
rs2777891	cg12054453	VMP1	2.97×10^{-20}
rs9355610	cg24160243	RPS6KA2	6.77×10^{-20}
rs933243	cg24160243	RPS6KA2	7.44×10^{-20}
rs11869228	cg12054453	VMP1	7.64×10^{-20}
rs2838738	cg04321224	ITGB2	1.00×10^{-19}
rs3782287	cg14849578	SCARB1	1.88×10^{-19}
rs9909324	cg12054453	VMP1	2.02×10^{-19}
rs9789054	cg12582317	MSI2	2.21×10^{-19}
rs989892	cg14849578	SCARB1	3.29×10^{-19}
rs7221815	cg12054453	VMP1	3.85×10^{-19}
rs10853015	cg16936953	VMP1	4.14×10^{-19}
rs3173833	cg07698196	TMEM176B	4.75×10^{-19}
rs3173833	cg07698196	TMEM176B	4.75×10^{-19}

Table A.7: The 100 most significant MeQTLs in the IBD Character cohort

A.4 Supplemental

The complete tables of significant results for IBD, CD, UC, and CD vs UC from the preliminary analysis of the IBD Character dataset are shown below.

MAF	Probe	Chr	Coord		Gene	LogFC	P Value
0.027	cg09349128	22	50327986	*	CRELD2	-0.25	8.41×10^{-21}
	cg12054453	17	57915717		VMP1	-0.55	3.66×10^{-20}
	cg17501210	6	166970252		RPS6KA2	-0.13	1.08×10^{-19}
	cg12170787	19	1130965		SBNO2	-0.13	2.70×10^{-19}
	cg18608055	19	1130866		SBNO2	-0.16	2.85×10^{-19}
	cg07573872	19	1126342		SBNO2	-0.15	2.22×10^{-18}
	cg16936953	17	57915665		VMP1	-0.32	4.05×10^{-18}
	cg27469606	19	1154485		SBNO2	-0.06	2.73×10^{-17}
	cg01059398	3	172235808		TNFSF10	-0.15	1.15×10^{-15}
	cg02650017	17	47301614		PHOSPHO1	-0.31	1.32×10^{-15}
	cg18513344	3	195531298		MUC4	-0.14	5.45×10^{-15}
	cg03156547	14	23525003		CDH24	-0.05	1.68×10^{-14}
	cg24002003	15	101668143	†	CHSY1	-0.14	1.76×10^{-14}
	cg02716826	9	33447032		AQP3	-0.17	1.95×10^{-14}
	cg06164260	3	187454439		BCL6	-0.08	2.91×10^{-14}
	cg13781414	9	138951648		NACC2	-0.06	8.55×10^{-14}
	cg10636246	1	159046973	*	AIM2	-0.21	1.74×10^{-13}
	cg18942579	17	57915773		VMP1	-0.22	1.75×10^{-13}
	cg01983725	1	227161799		ADCK3	-0.04	4.23×10^{-13}
	cg19821297	19	12890029	*	HOOK2	-0.16	5.70×10^{-13}
	cg16292768	8	27467783		CLU	-0.14	5.73×10^{-13}
	cg00159243	12	109023799		SELPLG	-0.18	6.96×10^{-13}
	cg02734358	4	90227074		GPRIN3	-0.20	7.27×10^{-13}
	cg13447080	8	131054408	*	ASAP1	-0.08	1.18×10^{-12}
	cg11393173	1	116369577	*	NHLH2	-0.10	1.29×10^{-12}
	cg15551881	9	123688715		TRAF1	0.26	1.68×10^{-12}
	cg01409343	17	57915740		VMP1	-0.16	1.77×10^{-12}
	cg03029755	17	76265444		LOC100996291	-0.07	3.21×10^{-12}
	cg24129923	17	7814251		CHD3	0.06	3.36×10^{-12}
	cg03998636	13	111210121		RAB20	-0.08	7.83×10^{-12}
	cg20995564	2	145172035		ZEB2	-0.25	1.05×10^{-11}
0.159	cg13028635	11	1871244	*	LSP1	0.05	1.06×10^{-11}

Table A.8: IBD Results: Continued overleaf

MAF	Probe	Chr	Coord		Gene	LogFC	P Value
	cg12269535	6	43142014		SRF	-0.09	1.26×10^{-11}
	cg10967866	10	134362164		INPP5A	0.09	2.22×10^{-11}
	cg12992827	3	101901234		ZPLD1	-0.17	2.60×10^{-11}
	cg19281794	3	112218761	*	BTLA	0.22	2.81×10^{-11}
0.012	cg13789303	6	108886681		FOXO3	0.12	3.02×10^{-11}
	cg21190595	11	3071167		CARS	-0.07	3.73×10^{-11}
	cg20962215	19	54713514	*	RPS9	-0.07	3.84×10^{-11}
	cg13165240	17	3715743		NCBP3	0.09	5.28×10^{-11}
	cg13442606	11	3071269		CARS	-0.05	6.29×10^{-11}
	cg02341197	21	34185927		C21orf62	0.13	6.94×10^{-11}
0.015	cg19283354	17	41851905		DUSP3	-0.06	7.06×10^{-11}
	cg12582317	17	55822272	†	MSI2	0.09	1.01×10^{-10}
	cg07793033	16	85256423	†	LINC00311	0.07	1.08×10^{-10}
	cg08449144	11	1857919		SYT8	-0.05	1.40×10^{-10}
0.500	cg02782634	17	57916643		VMP1	-0.14	1.44×10^{-10}
	cg13585930	10	72027357		NPFFR1	-0.15	1.45×10^{-10}
	cg23556574	10	13981154		FRMD4A	0.07	1.55×10^{-10}
	cg03940776	6	158490013		SYNJ2	-0.08	1.88×10^{-10}
	cg14849578	12	125282480		SCARB1	0.06	1.97×10^{-10}
	cg12680205	10	80303482		LINC00856	-0.07	2.23×10^{-10}
	cg18860310	4	87752504		SLC10A6	-0.10	3.00×10^{-10}
	cg08159663	16	57022486	*	NLRC5	-0.05	3.17×10^{-10}
	cg10515048	14	50697283		SOS2	0.19	3.45×10^{-10}
	cg15772366	6	111926749		TRAF3IP2	0.24	3.61×10^{-10}
	cg19137806	10	134362170		INPP5A	0.07	4.49×10^{-10}
	cg20692268	1	25358981	†	RUNX3	-0.13	4.80×10^{-10}
	cg22488164	12	14716910		PLBD1	0.08	4.94×10^{-10}
	cg04848343	19	4544095		SEMA6B	0.08	5.10×10^{-10}
	cg10472711	7	797592		DNAAF5	-0.08	5.29×10^{-10}
	cg13590277	5	150019496		SYNPO	-0.05	5.80×10^{-10}
	cg17260706	11	118782879	*	BCL9L	-0.09	6.41×10^{-10}
	cg04321224	21	46341380		ITGB2	0.15	7.83×10^{-10}
	cg18734877	8	27297419		PTK2B	-0.07	8.02×10^{-10}
	cg18663307	21	46341389		ITGB2	0.16	8.87×10^{-10}
	cg12370935	17	16976475		MPRIP	-0.06	9.53×10^{-10}
	cg09018739	16	57180107		CPNE2	0.07	1.06×10^{-9}

Table A.8: IBD Results: Continued overleaf

MAF	Probe	Chr	Coord		Gene	LogFC	P Value
	cg11859594	17	76755745		CYTH1	0.15	1.09×10^{-9}
	cg25739715	22	30663881	*	OSM	-0.17	1.09×10^{-9}
	cg26804423	7	8201134		ICA1	0.07	1.20×10^{-9}
	cg15981982	2	8685837	‡	ID2	-0.08	1.22×10^{-9}
	cg04324276	17	17817462		TOM1L2	-0.07	1.25×10^{-9}
	cg20748397	17	18062972		MYO15A	-0.04	1.33×10^{-9}
	cg05279866	8	37378355	‡	ZNF703	0.07	1.36×10^{-9}
	cg01145119	8	144441955		TOP1MT	0.06	1.45×10^{-9}
	cg15321908	3	53187213	*	PRKCD	-0.05	1.46×10^{-9}
	cg10460946	17	76247467	*	LOC100996291	-0.28	1.46×10^{-9}
	cg02464912	14	64319543	*	SYNE2	-0.22	1.50×10^{-9}
	cg13997435	1	153538406	*	S100A2	-0.05	1.58×10^{-9}
	cg00049440	9	73026643		KLF9	-0.19	1.70×10^{-9}
	cg05968822	1	226071684	*	TMEM63A	0.06	1.85×10^{-9}
	cg06946543	13	76377366		LMO7	0.17	2.03×10^{-9}
	cg17901584	1	55353706	*	DHCR24	-0.12	2.16×10^{-9}
	cg08289839	13	111318640		CARS2	0.05	2.35×10^{-9}
	cg00053916	8	37457329	†	ZNF703	0.04	2.59×10^{-9}
	cg23172671	1	203482523	*	OPTC	0.18	2.65×10^{-9}
	cg22356061	1	227954102		SNAP47	-0.10	2.89×10^{-9}
	cg04987734	14	103415873		CDC42BPB	0.11	2.96×10^{-9}
	cg14722693	8	19436451		CSGALNACT1	-0.08	3.05×10^{-9}
	cg12594635	12	54129457	*	CALCOCO1	-0.04	3.07×10^{-9}
	cg12669088	12	25541364	†	LMNTD1	-0.06	3.45×10^{-9}
	cg21773162	2	98330087		ZAP70	-0.05	3.65×10^{-9}
	cg20701457	17	62084342		ICAM2	-0.05	3.72×10^{-9}
	cg16755067	17	72259202	*	TTYH2	-0.04	3.88×10^{-9}
	cg18181703	17	76354621		SOCS3	-0.11	3.96×10^{-9}
	cg02538891	9	139549426	*	EGFL7	-0.05	3.96×10^{-9}
	cg26227957	1	19547285		EMC1	0.08	4.10×10^{-9}
	cg20141108	1	165907859	†	UCK2	0.09	4.25×10^{-9}
	cg16736826	1	41951512	*	EDN2	-0.06	4.43×10^{-9}
0.023	cg11415156	2	20779199	†	HS1BP3	-0.06	4.61×10^{-9}
	cg26470501	19	45252955		BCL3	-0.08	4.91×10^{-9}
	cg00490406	1	159046773	*	AIM2	-0.25	5.26×10^{-9}
	cg25368647	5	176736591		MXD3	-0.09	5.35×10^{-9}

Table A.8: IBD Results: Continued overleaf

MAF	Probe	Chr	Coord		Gene	LogFC	P Value
	cg15429134	12	12619103		BORCS5	0.15	5.71×10^{-9}
	cg20775254	2	95940705		PROM2	-0.04	5.78×10^{-9}
0.233	cg18991601	2	208635549	*	FZD5	0.19	5.82×10^{-9}
	cg05318454	19	46117805		EML2	-0.06	5.84×10^{-9}
	cg17803993	3	71276214		FOXP1	-0.14	6.43×10^{-9}
	cg03147185	2	97008030		NCAPH	-0.05	7.01×10^{-9}
	cg18363008	17	4675114	*	TM4SF5	-0.05	7.04×10^{-9}
	cg07698196	7	150485638	*	TMEM176B	-0.08	7.75×10^{-9}
	cg14989316	10	80757927		ZMIZ1-AS1	-0.06	8.24×10^{-9}
	cg24437104	1	227951794		SNAP47	-0.05	8.39×10^{-9}
	cg24215776	11	65152051	*	SLC25A45	-0.09	8.54×10^{-9}
	cg24536818	12	55371892		TESPA1	0.22	8.72×10^{-9}
	cg26158270	11	12309622		MICALCL	-0.06	8.80×10^{-9}
	cg23902264	1	241815413	*	WDR64	0.09	8.89×10^{-9}
	cg06383241	12	116997022	*	MAP1LC3B2	-0.20	9.01×10^{-9}
	cg20408175	12	53693169		PFDN5	-0.05	9.06×10^{-9}
	cg07267600	12	2750053		CACNA1C	-0.11	9.28×10^{-9}
	cg26663590	16	28959310	*	NFATC2IP	0.10	9.53×10^{-9}
	cg04310331	11	2012332	*	MRPL23-AS1	-0.08	9.74×10^{-9}
	cg25132241	14	92396859		FBLN5	0.04	1.05×10^{-8}
	cg21727145	15	101458127	*	ALDH1A3	0.07	1.05×10^{-8}
	cg08539067	3	49395985	*	GPX1	-0.18	1.07×10^{-8}
	cg02508743	8	56903623		LYN	0.14	1.13×10^{-8}
	cg15082870	7	36022841	†	SEPT7	0.08	1.19×10^{-8}
	cg25114611	6	35696870		LOC285847	-0.12	1.27×10^{-8}
	cg07120889	1	150535935	*	ADAMTSL4	-0.11	1.29×10^{-8}
	cg14303526	10	101609768		ABCC2	0.07	1.39×10^{-8}
	cg26680989	16	85560739	†	GSE1	0.06	1.47×10^{-8}
	cg25112191	1	151804260		RORC	-0.08	1.49×10^{-8}
	cg16809457	6	90399677		MDN1	0.08	1.53×10^{-8}
	cg04496906	8	145101182		SPATC1	-0.05	1.55×10^{-8}
	cg16505704	13	80378331	‡	NDFIP2	0.05	1.65×10^{-8}
	cg01209199	6	43274833		CRIP3	-0.03	1.67×10^{-8}
	cg11313780	8	103817950	†	AZIN1	-0.06	1.77×10^{-8}
	cg07683636	2	219940977		NHEJ1	0.04	1.81×10^{-8}
	cg18637901	4	3247372	*	HTT	-0.04	1.99×10^{-8}

Table A.8: IBD Results: Continued overleaf

MAF	Probe	Chr	Coord		Gene	LogFC	P Value
	cg10330371	7	122527499	*	CADPS2	0.15	2.05×10^{-8}
	cg10773266	10	43895225		HNRNPF	-0.07	2.17×10^{-8}
	cg12053291	12	125282342		SCARB1	0.04	2.42×10^{-8}
	cg12308965	11	67291387	*	CABP2	-0.04	2.43×10^{-8}
	cg01026383	6	32905012		HLA-DMB	0.03	2.48×10^{-8}
	cg07035242	1	11336263		UBIAD1	0.05	2.50×10^{-8}
	cg14114377	19	45756845		MARK4	-0.05	2.56×10^{-8}
	cg03875678	14	25103546	*	GZMB	-0.10	2.61×10^{-8}
	cg22562591	8	82002977		PAG1	0.15	2.70×10^{-8}
	cg02003183	14	103415882		CDC42BPB	0.23	2.71×10^{-8}
	cg07967630	6	58149317	†	GUSBP4	0.22	2.73×10^{-8}
	cg04095776	6	31106941		PSORS1C1	-0.04	2.77×10^{-8}
	cg14355374	10	3850160	†	KLF6	-0.05	2.82×10^{-8}
	cg08913523	8	126649807	‡	TRIB1	-0.16	2.86×10^{-8}
	cg26748477	17	38516415		RARA	-0.05	2.86×10^{-8}
	cg06724236	14	67853825		PLEK2	0.09	3.06×10^{-8}
	cg14269477	7	142630597		TRPV5	-0.02	3.30×10^{-8}
	cg17602444	7	644819		PRKAR1B	-0.08	3.42×10^{-8}
	cg20271330	1	204338515		LINC00628	-0.05	3.52×10^{-8}
	cg16757384	12	125030150		NCOR2	-0.08	3.56×10^{-8}
	cg12459502	18	60904237		BCL2	0.08	3.93×10^{-8}
	cg21518947	11	67222648		CABP4	-0.03	3.97×10^{-8}
	cg24073994	17	38511770		RARA	-0.03	3.99×10^{-8}
	cg20691205	13	36919674		SPG20	0.11	4.06×10^{-8}
	cg05820066	16	75145843		LDHD	-0.08	4.14×10^{-8}
	cg03845363	17	17753516		TOM1L2	-0.10	4.54×10^{-8}
	cg05316065	8	130799007		GSDMC	-0.07	4.62×10^{-8}
	cg18143317	11	2037124	†	H19	-0.04	4.74×10^{-8}
0.017	cg13150925	12	124529439		FAM101A	0.09	4.75×10^{-8}
	cg03432225	1	228294401		MRPL55	-0.03	4.90×10^{-8}
	cg15614653	2	129104464	†	HS6ST1	-0.04	4.93×10^{-8}
	cg05246110	11	68238943		PPP6R3	-0.04	4.94×10^{-8}
	cg17332198	12	58209913	*	AVIL	-0.07	4.98×10^{-8}
	cg25840433	6	167104903		RPS6KA2	-0.04	5.16×10^{-8}
	cg15050111	1	24645850		GRHL3	0.16	5.47×10^{-8}
	cg00705600	4	185820660	*	LINC01093	-0.05	5.82×10^{-8}

Table A.8: IBD Results: Continued overleaf

MAF	Probe	Chr	Coord		Gene	LogFC	P Value
	cg18131689	11	22214333	*	ANO5	0.10	5.83×10^{-8}
	cg23866916	19	1155738		SBNO2	-0.06	6.01×10^{-8}
	cg05937055	1	181128764	†	IER5	-0.10	6.46×10^{-8}
	cg05946309	16	85926085	*	IRF8	0.03	6.82×10^{-8}
	cg18396675	17	56605391		SEPT4	-0.03	6.82×10^{-8}
0.017	cg04389058	3	57041402		ARHGEF3	-0.06	7.14×10^{-8}
	cg21337881	17	5138645		LOC100130950	0.06	7.27×10^{-8}
	cg00009293	7	44151172		AEBP1	-0.03	7.31×10^{-8}
	cg12616487	11	62379063		EML3	-0.04	7.86×10^{-8}
	cg14965639	2	48795994		STON1-GTF2A1L	-0.17	8.71×10^{-8}
	cg00908292	17	38699009	†	CCR7	-0.07	8.81×10^{-8}
	cg02345399	11	65195039		NEAT1	-0.07	8.88×10^{-8}
0.01	cg07539709	17	80545454		FOXK2	0.10	8.89×10^{-8}
	cg00107490	22	19509318	*	CDC45	-0.02	9.56×10^{-8}
	cg24637035	12	120445227		CCDC64	0.11	9.58×10^{-8}
	cg11267810	20	19867026	*	RIN2	0.03	9.75×10^{-8}
	cg12687767	20	19870109		RIN2	-0.05	1.01×10^{-7}

Table A.8: Illumina 450k probes with Bonferroni significant IBD-associated methylation changes in the IBD Character cohort. Where the probe is not in a gene the nearest gene is given with an indication of distance: *<10kb, †<100kb, ‡>100kb. Where the is a SNP within 10bp the MAF (or highest if there are multiple) is shown on the left.

MAF	Probe	Chr	Coord		Gene	LogFC	P Value
	cg12170787	19	1130965		SBNO2	-0.16	1.45×10^{-22}
	cg18608055	19	1130866		SBNO2	-0.20	5.52×10^{-21}
0.027	cg09349128	22	50327986	*	CRELD2	-0.28	1.13×10^{-18}
	cg12054453	17	57915717		VMP1	-0.63	1.58×10^{-17}
	cg17501210	6	166970252		RPS6KA2	-0.14	4.26×10^{-17}
	cg07573872	19	1126342		SBNO2	-0.16	9.54×10^{-17}
	cg16936953	17	57915665		VMP1	-0.32	1.72×10^{-14}
	cg27469606	19	1154485		SBNO2	-0.06	2.21×10^{-14}
	cg01059398	3	172235808		TNFSF10	-0.16	3.12×10^{-12}
	cg10967866	10	134362164		INPP5A	0.11	3.18×10^{-12}
	cg18513344	3	195531298		MUC4	-0.14	4.29×10^{-12}
	cg15551881	9	123688715		TRAF1	0.31	8.30×10^{-12}
	cg20141108	1	165907859	†	UCK2	0.12	9.63×10^{-12}
	cg19821297	19	12890029	*	HOOK2	-0.17	1.04×10^{-11}
	cg13781414	9	138951648		NACC2	-0.06	1.23×10^{-11}
	cg18181703	17	76354621		SOCS3	-0.14	1.35×10^{-11}
	cg02650017	17	47301614		PHOSPHO1	-0.32	2.14×10^{-11}
	cg16292768	8	27467783		CLU	-0.15	3.51×10^{-11}
	cg18942579	17	57915773		VMP1	-0.23	6.68×10^{-11}
	cg03940776	6	158490013		SYNJ2	-0.09	1.07×10^{-10}
	cg26804423	7	8201134		ICA1	0.08	1.31×10^{-10}
	cg19137806	10	134362170		INPP5A	0.09	1.54×10^{-10}
0.027	cg19281794	3	112218761	*	BTLA	0.25	2.25×10^{-10}
	cg01409343	17	57915740		VMP1	-0.17	2.49×10^{-10}
	cg20995564	2	145172035		ZEB2	-0.25	2.93×10^{-10}
	cg20701457	17	62084342		ICAM2	-0.07	4.93×10^{-10}
	cg01145119	8	144441955		TOP1MT	0.07	5.00×10^{-10}
	cg00159243	12	109023799		SELPLG	-0.19	5.63×10^{-10}
	cg23172671	1	203482523	*	OPTC	0.23	1.19×10^{-9}
	cg25112191	1	151804260		RORC	-0.10	1.49×10^{-9}
	cg12269535	6	43142014		SRF	-0.09	2.98×10^{-9}
	cg13573513	12	14996143		ART4	0.11	3.19×10^{-9}
	cg15082870	7	36022841	†	SEPT7	0.10	3.53×10^{-9}
	cg06164260	3	187454439		BCL6	-0.07	3.90×10^{-9}
	cg24002003	15	101668143	†	CHSY1	-0.13	4.74×10^{-9}
	cg26340740	11	58980377		MPEG1	0.18	5.16×10^{-9}

Table A.9: CD Results: Continued overleaf

MAF	Probe	Chr	Coord		Gene	LogFC	P Value
	cg12535090	11	20071677		NAV2	0.05	6.13×10^{-9}
	cg01983725	1	227161799		ADCK3	-0.04	7.78×10^{-9}
	cg21727145	15	101458127	*	ALDH1A3	0.08	7.82×10^{-9}
	cg09018739	16	57180107		CPNE2	0.08	7.87×10^{-9}
0.500	cg02782634	17	57916643		VMP1	-0.14	8.70×10^{-9}
	cg26227957	1	19547285		EMC1	0.09	1.06×10^{-8}
	cg19695507	10	13526193		BEND7	0.20	1.07×10^{-8}
	cg03029755	17	76265444		LOC100996291	-0.07	1.08×10^{-8}
	cg08363114	19	54711292		RPS9	-0.14	1.19×10^{-8}
0.159	cg13028635	11	1871244	*	LSP1	0.06	1.25×10^{-8}
	cg03998636	13	111210121		RAB20	-0.07	1.40×10^{-8}
	cg19769147	14	105860954		PACS2	0.06	1.40×10^{-8}
	cg16395997	1	3562798		WRAP73	0.11	1.47×10^{-8}
	cg12992827	3	101901234		ZPLD1	-0.15	1.96×10^{-8}
	cg05272827	17	79380706		BAHCC1	0.13	2.07×10^{-8}
	cg00705600	4	185820660	*	LINC01093	-0.05	2.20×10^{-8}
	cg16809457	6	90399677		MDN1	0.09	2.23×10^{-8}
	cg01026383	6	32905012		HLA-DMB	0.04	2.64×10^{-8}
0.012	cg11551560	15	70528789	‡	TLE3	0.08	2.92×10^{-8}
	cg08289839	13	111318640		CARS2	0.06	2.96×10^{-8}
0.196	cg15599437	5	40676198	*	PTGER4	0.29	3.06×10^{-8}
	cg22488164	12	14716910		PLBD1	0.09	3.76×10^{-8}
	cg02716826	9	33447032		AQP3	-0.15	4.03×10^{-8}
	cg14989316	10	80757927		ZMIZ1-AS1	-0.07	4.12×10^{-8}
	cg13139542	2	8242815		LINC00299	0.09	4.77×10^{-8}
	cg06724236	14	67853825		PLEK2	0.11	5.90×10^{-8}
	cg02341197	21	34185927		C21orf62	0.13	6.44×10^{-8}
	cg10508317	17	76355146		SOCS3	-0.26	6.66×10^{-8}
	cg07793033	16	85256423	†	LINC00311	0.07	6.73×10^{-8}
	cg17901584	1	55353706	*	DHCR24	-0.13	6.77×10^{-8}
	cg10460946	17	76247467	*	LOC100996291	-0.30	6.78×10^{-8}
	cg05279866	8	37378355	‡	ZNF703	0.08	7.25×10^{-8}
	cg17953136	2	20232577		LAPTM4A	0.12	7.30×10^{-8}
	cg07298772	2	176306492	‡	ATP5G3	0.22	7.45×10^{-8}
	cg10472711	7	797592		DNAAF5	-0.08	7.74×10^{-8}
	cg23866916	19	1155738		SBNO2	-0.07	7.90×10^{-8}

Table A.9: CD Results: Continued overleaf

MAF	Probe	Chr	Coord		Gene	LogFC	P Value
	cg23556574	10	13981154		FRMD4A	0.08	8.43×10^{-8}
	cg04321224	21	46341380		ITGB2	0.16	8.48×10^{-8}
	cg09287328	10	134231487	*	PWWP2B	0.12	8.56×10^{-8}
	cg26663590	16	28959310	*	NFATC2IP	0.11	8.74×10^{-8}
	cg18663307	21	46341389		ITGB2	0.18	8.78×10^{-8}
	cg15614653	2	129104464	†	HS6ST1	-0.05	9.07×10^{-8}
	cg10636246	1	159046973	*	AIM2	-0.19	9.75×10^{-8}
	cg17330097	1	206316676	*	CTSE	0.08	9.03×10^{-8}
	cg18860310	4	87752504		SLC10A6	-0.10	1.02×10^{-7}

Table A.9: Illumina 450k probes with Bonferroni significant CD-associated methylation changes in the IBD Character cohort. Where the probe is not in a gene the nearest gene is given with an indication of distance: *<10kb, †<100kb, ‡>100kb. Where the is a SNP within 10bp the MAF (or highest if there are multiple) is shown on the left.

MAF	Probe	Chr	Coord		Gene	LogFC	P Value
0.027	cg09349128	22	50327986	*	CRELD2	-0.26	1.32×10^{-17}
	cg16936953	17	57915665		VMP1	-0.34	1.50×10^{-15}
	cg12054453	17	57915717		VMP1	-0.55	4.01×10^{-15}
	cg17501210	6	166970252		RPS6KA2	-0.13	6.74×10^{-15}
	cg07573872	19	1126342		SBNO2	-0.16	8.53×10^{-15}
	cg24002003	15	101668143	†	CHSY1	-0.17	1.16×10^{-14}
	cg18608055	19	1130866		SBNO2	-0.15	7.02×10^{-14}
	cg08449144	11	1857919		SYT8	-0.07	9.23×10^{-14}
	cg06164260	3	187454439		BCL6	-0.09	1.09×10^{-13}
	cg02650017	17	47301614		PHOSPHO1	-0.34	1.59×10^{-13}
	cg27469606	19	1154485		SBNO2	-0.06	2.02×10^{-13}
	cg01059398	3	172235808		TNFSF10	-0.17	2.68×10^{-13}
	cg02716826	9	33447032		AQP3	-0.19	7.34×10^{-13}
	cg10636246	1	159046973	*	AIM2	-0.25	1.56×10^{-12}
	cg01983725	1	227161799		ADCK3	-0.04	1.77×10^{-12}
	cg03156547	14	23525003		CDH24	-0.05	2.38×10^{-12}
	cg20995564	2	145172035		ZEB2	-0.29	3.54×10^{-12}
	cg03039243	1	2782699	†	TTC34	-0.07	5.87×10^{-12}
	cg02734358	4	90227074		GPRIN3	-0.24	6.26×10^{-12}
	cg12170787	19	1130965		SBNO2	-0.11	6.69×10^{-12}
	cg18513344	3	195531298		MUC4	-0.14	6.81×10^{-11}
	cg11393173	1	116369577	*	NHLH2	-0.11	7.56×10^{-12}
	cg18942579	17	57915773		VMP1	-0.24	9.39×10^{-12}
	cg13442606	11	3071269		CARS	-0.07	1.04×10^{-11}
	cg24129923	17	7814251		CHD3	0.07	1.45×10^{-11}
	cg19821297	19	12890029	*	HOOK2	-0.17	1.59×10^{-11}
	cg13447080	8	131054408	*	ASAP1	-0.09	1.68×10^{-11}
	cg12680205	10	80303482		LINC00856	-0.08	1.73×10^{-11}
	cg12992827	3	101901234		ZPLD1	-0.20	1.75×10^{-11}
	cg14849578	12	125282480		SCARB1	0.08	1.84×10^{-11}
	cg15981982	2	8685837	‡	ID2	-0.10	2.07×10^{-11}
	cg15321908	3	53187213	*	PRKCD	-0.07	2.23×10^{-11}
	cg25840433	6	167104903		RPS6KA2	-0.07	2.33×10^{-11}
	cg03998636	13	111210121		RAB20	-0.09	2.72×10^{-11}
	cg21190595	11	3071167		CARS	-0.08	3.25×10^{-11}
	cg00159243	12	109023799		SELPLG	-0.20	3.87×10^{-11}

Table A.10: UC Results: Continued overleaf

MAF	Probe	Chr	Coord		Gene	LogFC	P Value
	cg04436528	8	143554875		ADGRB1	-0.05	4.22×10^{-11}
	cg18860310	4	87752504		SLC10A6	-0.13	4.69×10^{-11}
0.015	cg19283354	17	41851905		DUSP3	-0.08	4.79×10^{-11}
	cg13781414	9	138951648		NACC2	-0.06	1.72×10^{-10}
	cg22488164	12	14716910		PLBD1	0.10	7.29×10^{-10}
	cg16292768	8	27467783		CLU	-0.14	6.17×10^{-10}
	cg13165240	17	3715743		NCBP3	0.11	1.10×10^{-10}
	cg03029755	17	76265444		LOC100996291	-0.08	5.50×10^{-10}
	cg10472711	7	797592		DNAAF5	-0.10	2.30×10^{-10}
	cg13028635	11	1871244	*	LSP1	0.06	4.06×10^{-10}
	cg02341197	21	34185927		C21orf62	0.14	3.53×10^{-10}
	cg20962215	19	54713514	*	RPS9	-0.07	2.87×10^{-10}
	cg12669088	12	25541364	†	LMNTD1	-0.08	6.50×10^{-10}
	cg26227957	1	19547285		EMC1	0.10	5.60×10^{-10}
	cg05318454	19	46117805		EML2	-0.07	2.56×10^{-10}
	cg03147185	2	97008030		NCAPH	-0.07	2.61×10^{-10}
	cg18363008	17	4675114	*	TM4SF5	-0.07	1.64×10^{-10}
	cg06383241	12	116997022	*	MAP1LC3B2	-0.26	2.92×10^{-10}
	cg20408175	12	53693169		PFDN5	-0.07	2.24×10^{-10}
	cg08913523	8	126649807	‡	TRIB1	-0.21	9.96×10^{-10}
	cg15050111	1	24645850		GRHL3	0.23	1.55×10^{-10}
	cg15557758	1	13828828		LRRC38	-0.04	5.20×10^{-10}
	cg18396675	17	56605391		SEPT4	-0.05	6.31×10^{-10}
	cg09803002	11	119039103	*	NLRX1	-0.04	2.68×10^{-10}
	cg04782146	16	85112370		KIAA0513	0.07	5.56×10^{-10}
	cg07168939	8	143763412		PSCA	-0.08	6.01×10^{-10}
	cg08463231	15	78103228	‡	LOC645752	-0.05	1.01×10^{-9}
	cg11986385	11	19628828		NAV2	-0.05	1.11×10^{-9}
	cg05007163	22	39266069		CBX6	-0.05	1.11×10^{-9}
	cg04299274	17	25798741	*	KSR1	-0.05	1.13×10^{-9}
	cg10330371	7	122527499	*	CADPS2	0.20	1.16×10^{-9}
	cg09419102	11	65550444	*	AP5B1	-0.05	1.24×10^{-9}
	cg10773266	10	43895225		HNRNPF	-0.10	1.25×10^{-9}
	cg20692268	1	25358981	†	RUNX3	-0.15	1.32×10^{-9}
	cg18583702	2	3700907	*	ALLC	-0.04	1.43×10^{-9}
	cg16736826	1	41951512	*	EDN2	-0.08	1.46×10^{-9}

Table A.10: UC Results: Continued overleaf

MAF	Probe	Chr	Coord		Gene	LogFC	P Value
	cg00193668	9	35814283		HINT2	-0.09	1.53×10^{-9}
	cg26325174	14	101360687	*	MEG8	-0.05	1.57×10^{-9}
	cg20691205	13	36919674		SPG20	0.15	1.79×10^{-9}
	cg01409343	17	57915740		VMP1	-0.16	1.81×10^{-9}
	cg24926042	12	52799999		KRT82	-0.04	1.87×10^{-9}
	cg26748477	17	38516415		RARA	-0.07	1.97×10^{-9}
	cg03179542	7	2647333		IQCE	-0.07	2.75×10^{-9}
	cg27115863	22	37921640	*	CARD10	-0.16	2.79×10^{-9}
	cg26501007	8	126614261	‡	TRIB1	-0.09	2.88×10^{-9}
	cg20748397	17	18062972		MYO15A	-0.05	2.94×10^{-9}
	cg12594635	12	54129457	*	CALCOCO1	-0.05	2.95×10^{-9}
	cg00014260	8	10383071		PRSS55	-0.05	3.10×10^{-9}
	cg08539067	3	49395985	*	GPX1	-0.22	3.15×10^{-9}
	cg15772366	6	111926749		TRAF3IP2	0.28	3.23×10^{-9}
	cg10967866	10	134362164		INPP5A	0.09	3.28×10^{-9}
	cg03241649	19	44405924		LOC100505715	0.15	3.46×10^{-9}
	cg05316065	8	130799007		GSDMC	-0.09	3.54×10^{-9}
	cg04321224	21	46341380		ITGB2	0.17	3.81×10^{-9}
	cg12616487	11	62379063		EML3	-0.06	4.21×10^{-9}
	cg18663307	21	46341389		ITGB2	0.19	4.27×10^{-9}
	cg14722693	8	19436451		CSGALNACT1	-0.09	4.28×10^{-9}
	cg12682367	13	24523698	*	ANKRD20A19P	-0.05	4.44×10^{-9}
	cg04496906	8	145101182		SPATC1	-0.07	4.59×10^{-9}
	cg13585930	10	72027357		NPFFR1	-0.16	4.72×10^{-9}
	cg22788465	5	131398845		IL3	-0.07	4.84×10^{-9}
	cg18506744	19	1102627	*	GPX4	-0.17	5.00×10^{-9}
	cg20711030	7	150897624	*	ABCF2	-0.05	5.19×10^{-9}
	cg13590277	5	150019496		SYNPO	-0.06	5.53×10^{-9}
	cg15647725	5	1091936		SLC12A7	-0.04	5.55×10^{-9}
	cg26325867	6	41398693	†	NCR2	-0.07	5.66×10^{-9}
	cg21448513	8	57103638		PLAG1	-0.06	5.71×10^{-9}
	cg27343325	20	3649247		ADAM33	-0.04	6.25×10^{-9}
	cg05824755	1	120613447	*	NOTCH2	-0.04	6.26×10^{-9}
	cg05279866	8	37378355	‡	ZNF703	0.08	6.30×10^{-9}
	cg23662734	17	16386804		LRRC75A	-0.04	6.40×10^{-9}
	cg25739715	22	30663881	*	OSM	-0.19	6.43×10^{-9}

Table A.10: UC Results: Continued overleaf

MAF	Probe	Chr	Coord		Gene	LogFC	P Value
	cg01209199	6	43274833		CRIP3	-0.04	6.59×10^{-9}
	cg17433678	10	3135764		PFKP	-0.15	6.86×10^{-9}
	cg00490406	1	159046773	*	AIM2	-0.31	6.88×10^{-9}
	cg04563718	1	2565777	*	MMEL1	-0.05	7.03×10^{-9}
	cg02464912	14	64319543	*	SYNE2	-0.26	7.72×10^{-9}
	cg04742397	14	95986554	†	SNHG10	-0.12	8.20×10^{-9}
	cg10241823	17	841205		NXN	0.09	8.67×10^{-9}
	cg12274082	6	32007986		CYP21A2	-0.04	8.81×10^{-9}
	cg07035242	1	11336263		UBIAD1	0.06	9.11×10^{-9}
	cg19054096	20	59868424		CDH4	-0.04	9.18×10^{-9}
	cg14355374	10	3850160	†	KLF6	-0.06	9.18×10^{-9}
	cg17206420	2	27482803		SLC30A3	-0.06	9.22×10^{-9}
	cg00705600	4	185820660	*	LINC01093	-0.06	1.00×10^{-8}
	cg14269477	7	142630597		TRPV5	-0.03	1.04×10^{-8}
	cg20116935	3	50303757	*	SEMA3B	-0.06	1.06×10^{-8}
	cg04567334	10	73408185		CDH23	-0.04	1.11×10^{-8}
	cg01514353	22	46259643	†	ATXN10	-0.16	1.13×10^{-8}
	cg13256912	3	11211081		HRH1	-0.06	1.13×10^{-8}
	cg05304729	1	158800024	*	MNDA	0.25	1.20×10^{-8}
	cg01668099	11	130026798	*	ST14	-0.04	1.21×10^{-8}
	cg24157892	17	59473707	*	BCAS3	0.27	1.22×10^{-8}
	cg07914457	1	27241853	*	NR0B2	-0.05	1.23×10^{-8}
	cg08782002	17	61847885		CCDC47	-0.08	1.25×10^{-8}
	cg18143317	11	2037124	†	H19	-0.05	1.26×10^{-8}
	cg04987734	14	103415873		CDC42BPB	0.13	1.35×10^{-8}
	cg16008966	1	114761794	†	SYT6	-0.08	1.38×10^{-8}
	cg15727249	11	116693415		APOA4	-0.04	1.39×10^{-8}
0.061	cg09054949	4	74735101	*	CXCL1	0.27	1.40×10^{-8}
	cg22471129	22	30868226	*	SEC14L3	-0.04	1.43×10^{-8}
	cg23902264	1	241815413	*	WDR64	0.11	1.50×10^{-8}
	cg24136205	13	100624293	*	ZIC5	0.22	1.50×10^{-8}
	cg19281794	3	112218761	*	BTLA	0.23	1.52×10^{-8}
	cg15460584	4	8540687	†	GPR78	-0.06	1.72×10^{-8}
	cg22356061	1	227954102		SNAP47	-0.12	1.73×10^{-8}
	cg05095591	3	147129184		ZIC1	0.24	1.75×10^{-8}
	cg07267600	12	2750053		CACNA1C	-0.13	1.83×10^{-8}

Table A.10: UC Results: Continued overleaf

MAF	Probe	Chr	Coord		Gene	LogFC	P Value
	cg20510033	1	8960134	†	ENO1-AS1	-0.12	1.83×10^{-8}
	cg15114651	19	47289410		SLC1A5	-0.05	1.84×10^{-8}
	cg17260706	11	118782879	*	BCL9L	-0.10	1.89×10^{-8}
	cg13997435	1	153538406	*	S100A2	-0.05	1.96×10^{-8}
	cg21518947	11	67222648		CABP4	-0.04	1.96×10^{-8}
	cg03990033	1	208084030		CD34	0.21	2.01×10^{-8}
	cg21727145	15	101458127	*	ALDH1A3	0.08	2.06×10^{-8}
	cg08035323	2	9843525	†	YWHAQ	-0.21	2.06×10^{-8}
	cg07967630	6	58149317	†	GUSBP4	0.28	2.09×10^{-8}
	cg18734877	8	27297419		PTK2B	-0.08	2.21×10^{-8}
	cg04913803	14	77510549	†	IRF2BPL	-0.07	2.25×10^{-8}
	cg09018739	16	57180107		CPNE2	0.07	2.39×10^{-8}
	cg23432368	2	47882601		MSH2	-0.04	2.46×10^{-8}
	cg21422361	7	1884392		MAD1L1	-0.10	2.54×10^{-8}
	cg12756396	1	50886969		DMRTA2	0.19	2.57×10^{-8}
	cg08289839	13	111318640		CARS2	0.06	2.59×10^{-8}
	cg24226371	15	75945038		SNX33	-0.05	2.71×10^{-8}
	cg06206957	7	33080615		NT5C3A	0.16	2.72×10^{-8}
	cg26575914	1	228397977		C1orf145	-0.04	2.73×10^{-8}
	cg05209330	5	131550942		P4HA2	-0.04	2.84×10^{-8}
	cg20631044	19	41255890		C19orf54	-0.06	2.86×10^{-8}
	cg05968822	1	226071684	*	TMEM63A	0.07	2.87×10^{-8}
	cg26708319	4	111552040		PITX2	0.18	2.89×10^{-8}
	cg04465154	8	9045558	†	PPP1R3B	-0.14	2.90×10^{-8}
	cg06621080	1	201631166		NAV1	-0.06	2.91×10^{-8}
	cg05223210	11	117630397		DSCAML1	-0.05	2.95×10^{-8}
	cg04848343	19	4544095		SEMA6B	0.09	3.22×10^{-8}
	cg16298481	1	89989717	*	LRRC8B	-0.24	3.26×10^{-8}
	cg25114611	6	35696870		LOC285847	-0.13	3.27×10^{-8}
	cg12582317	17	55822272	†	MSI2	0.09	3.31×10^{-8}
	cg14058848	9	130479114		PTRH1	-0.06	3.37×10^{-8}
	cg04070847	3	112693867		CD200R1	0.16	3.40×10^{-8}
	cg00795341	5	79330929	*	THBS4	0.06	3.46×10^{-8}
	cg22320365	17	36718198		SRCIN1	0.30	3.69×10^{-8}
	cg03950000	1	160617327	*	SLAMF1	-0.06	3.70×10^{-8}
	cg21691893	6	32018110		TNXB	-0.05	3.70×10^{-8}

Table A.10: UC Results: Continued overleaf

MAF	Probe	Chr	Coord		Gene	LogFC	P Value
	cg17290213	11	65344355		EHBP1L1	-0.08	3.75×10^{-8}
	cg21241889	9	123690658		TRAF1	-0.12	3.83×10^{-8}
	cg06422108	10	31610399		ZEB1	0.26	3.84×10^{-8}
	cg20695203	13	112335939	‡	TEX29	-0.04	3.85×10^{-8}
	cg02325250	5	131409289	*	CSF2	-0.06	3.91×10^{-8}
	cg23712342	10	60936794		PHYHIPL	0.26	3.95×10^{-8}
0.012	cg13789303	6	108886681		FOXO3	0.11	4.11×10^{-8}
	cg08176258	7	1938450		MAD1L1	-0.09	4.15×10^{-8}
	cg08530237	12	22094563	*	ABCC9	0.22	4.16×10^{-8}
	cg18637901	4	3247372	*	HTT	-0.04	4.19×10^{-8}
	cg03875678	14	25103546	*	GZMB	-0.11	4.23×10^{-8}
	cg03519967	1	26010346		MAN1C1	-0.05	4.25×10^{-8}
0.500	cg23523668	11	65194989		NEAT1	-0.22	4.26×10^{-8}
	cg24882551	4	85403032	†	NKX6-1	0.18	4.29×10^{-8}
	cg04615865	12	121678720		CAMKK2	0.29	4.38×10^{-8}
	cg01436487	7	1890148		MAD1L1	-0.07	4.41×10^{-8}
	cg08386009	6	32094845		ATF6B	-0.05	4.43×10^{-8}
	cg27294813	2	109791599		SH3RF3	-0.05	4.45×10^{-8}
	cg24621042	14	94857275	*	SERPINA1	-0.10	4.55×10^{-8}
	cg10515048	14	50697283		SOS2	0.20	4.62×10^{-8}
	cg18110483	5	79330959	*	THBS4	0.23	4.67×10^{-8}
	cg09791504	1	151701849		RIIAD1	-0.07	4.96×10^{-8}
	cg03845363	17	17753516		TOM1L2	-0.12	5.12×10^{-8}
	cg04761198	8	54793733		RGS20	0.13	5.15×10^{-8}
	cg12483476	5	178421711		GRM6	0.46	5.18×10^{-8}
	cg20466954	1	84768359		SAMD13	0.11	5.31×10^{-8}
	cg03432225	1	228294401		MRPL55	-0.03	5.35×10^{-8}
	cg02743674	11	2433468		TRPM5	-0.04	5.46×10^{-8}
	cg04100532	2	210636560	*	UNC80	0.19	5.58×10^{-8}
	cg14114377	19	45756845		MARK4	-0.06	5.62×10^{-8}
	cg03465179	14	106054474	†	TMEM121	-0.05	5.71×10^{-8}
	cg21773162	2	98330087		ZAP70	-0.05	5.73×10^{-8}
	cg25368647	5	176736591		MXD3	-0.10	5.90×10^{-8}
	cg01165560	11	49911386	†	OR4C13	-0.04	6.02×10^{-8}
	cg06588876	9	139009275		C9orf69	-0.04	6.18×10^{-8}
	cg03699074	16	88849875		PIEZO1	-0.23	6.30×10^{-8}

Table A.10: UC Results: Continued overleaf

MAF	Probe	Chr	Coord		Gene	LogFC	P Value
	cg26562462	4	6929045		TBC1D14	-0.08	6.36×10^{-8}
0.500	cg03850035	2	222367110		EPHA4	-0.05	6.36×10^{-8}
	cg02629976	2	220411924		TMEM198	-0.05	6.41×10^{-8}
	cg11859594	17	76755745		CYTH1	0.16	6.43×10^{-8}
	cg10754002	2	121276805	†	LINC01101	-0.06	6.49×10^{-8}
	cg05937055	1	181128764	†	IER5	-0.12	6.77×10^{-8}
	cg13823169	9	139776893		TRAF2	-0.08	6.89×10^{-8}
	cg25529535	10	6125715	*	RBM17	-0.06	6.92×10^{-8}
0.032	cg17959631	7	44189684		GCK	-0.04	7.05×10^{-8}
	cg00084202	8	143360290		TSNARE1	-0.04	7.20×10^{-8}
	cg10734581	14	101520078	*	MIR382	-0.04	7.36×10^{-8}
	cg17280514	2	242755302		NEU4	-0.07	7.37×10^{-8}
	cg14646983	9	34602974	*	RPP25L	-0.05	7.54×10^{-8}
	cg23363971	12	56415271		IKZF4	0.17	7.75×10^{-8}
	cg04095776	6	31106941		PSORS1C1	-0.04	8.01×10^{-8}
	cg20673075	8	99838488		STK3	-0.08	8.02×10^{-8}
	cg17515347	1	159047163	*	AIM2	-0.29	8.06×10^{-8}
	cg19303305	4	8277560		HTRA3	-0.04	8.09×10^{-8}
	cg18346531	17	9967431		GAS7	-0.07	8.18×10^{-8}
	cg21895526	18	24765490		AQP4-AS1	0.23	8.23×10^{-8}
0.011	cg24661168	1	210013268		DIEXF	-0.07	8.32×10^{-8}
	cg09481056	16	2339260		ABCA3	0.05	8.37×10^{-8}
	cg06401614	11	44927976		TSPAN18	-0.04	8.58×10^{-8}
	cg07564837	8	58126962	*	LINC01606	-0.05	8.78×10^{-8}
	cg23986143	7	76054256		ZP3	-0.03	8.86×10^{-8}
	cg23689615	15	74228839		LOXL1	-0.06	8.87×10^{-8}
	cg05004855	5	149644009		CAMK2A	-0.10	9.01×10^{-8}
	cg07541160	6	36985879		FGD2	-0.04	9.37×10^{-8}
	cg02345399	11	65195039		NEAT1	-0.09	9.53×10^{-8}
	cg17585031	17	25798942	*	KSR1	-0.04	9.53×10^{-8}
	cg08858220	6	36805317		CPNE5	-0.04	9.61×10^{-8}
	cg05915362	6	159274480		OSTCP1	-0.12	9.72×10^{-8}
	cg15000071	3	134889493		EPHB1	-0.05	9.72×10^{-8}
	cg12308965	11	67291387	*	CABP2	-0.04	9.74×10^{-8}
	cg20964856	19	41767669		AXL	-0.08	9.78×10^{-8}
	cg13150925	12	124529439		FAM101A	0.11	9.91×10^{-8}

Table A.10: UC Results: Continued overleaf

MAF	Probe	Chr	Coord		Gene	LogFC	P Value
	cg12159215	11	118826779	*	UPK2	-0.06	9.92×10^{-8}
	cg04324276	17	17817462		TOM1L2	-0.08	1.00×10^{-7}
	cg05850205	8	67350187		ADHFE1	-0.04	1.00×10^{-7}
	cg16086570	5	141924566	†	FGF1	-0.04	1.01×10^{-7}
	cg12569119	10	105541150		SH3PXD2A	-0.04	1.02×10^{-7}
	cg07120889	1	150535935	*	ADAMTSL4	-0.13	1.02×10^{-7}
	cg14699903	3	48601227	*	UCN2	-0.04	1.03×10^{-7}

Table A.10: Illumina 450k probes with Bonferroni significant UC-associated methylation changes in the IBD Character cohort. Where the probe is not in a gene the nearest gene is given with an indication of distance: *<10kb, †<100kb, ‡>100kb. Where the is a SNP within 10bp the MAF (or highest if there are multiple) is shown on the left.

MAF	Probe	Chr	Coord	Gene	LogFC	P Value
	cg11986385	11	19628828	NAV2	0.05	6.82×10^{-8}

Table A.11: The single Bonferroni significant difference between CD and UC in the IBD Character cohort

Appendix B

Abstracts

B.1 BSG abstract

Genome-wide epigenetic analysis in childhood-onset Crohn's Disease implicates MIR21 [2]

[1]A T Adams, [1]N A Kennedy, [2]R Hansen, [1]N T Ventham, [1]K R O'Leary, [1]H E Drummond, [1]C L Noble, [2]E El-Omar, [3]R K Russell, [4]D C Wilson, [1]E R Nimmo, [2]G L Hold, [1]J Satsangi

[1]Gastrointestinal Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, UK. [2]Gastrointestinal Research Group, Division of Applied Medicine, University of Aberdeen, UK. [3]Department of Paediatric Gastroenterology, Royal Hospital for Sick Children, Glasgow, UK. [4]Paediatric Gastroenterology and Nutrition, Child Life and Health, University of Edinburgh, Royal Hospital for Sick Children, Edinburgh, UK.

Introduction: DNA methylation influences transcriptional activity and marks sites of active transcription. Technological developments allow the rapid assessment of methylation state at 450,000 sites across the genome. The aim of this study was to identify genes with a possible role in Crohn's disease pathogenesis, and candidate genes for methylation based diagnostic biomarkers.

Methods: Using the Illumina 450k platform we analysed genome-wide DNA methylation in symptomatic children who underwent diagnostic colonoscopy, half of whom were diagnosed with CD and half had no pathology. Replication was performed in children with established CD vs. symptomatic non-disease controls.

Further targeted replication by pyrosequencing was performed in adults with CD vs. healthy controls, with qPCR and microarray data to analyse expression.

Results: Meta-analysis of the combined paediatric datasets (n=66) identified 165 individual CpGs with epigenome-wide significance (Bonferroni correction) and 138 differentially methylated regions (DMR). Methylation changes were significantly enriched ($p < 0.0001$) in proximity to loci implicated by genome-wide association studies (GWAS). The strongest result by each approach was MicroRNA 21, within the autophagy gene VMP1 ($p = 1.2 \times 10^{-14}$), 48kb from GWAS SNP rs1292053. In adults with CD we replicated MIR21 hypomethylation ($p = 6.6 \times 10^{-5}$, $n = 172$), and showed increased expression in blood ($p < 0.005$, $n = 66$). Intestinal expression increased with inflammation in CD ($p = 1.4 \times 10^{-6}$, $n = 99$) but not controls ($n = 73$). Linear discriminant analysis of methylation in the paediatric discovery cohort accurately predicted disease state in the paediatric replication cohort (94% sensitivity, 100% specificity) based on methylation at two CpG sites.

Conclusion: MIR21 emerges as a target for further investigation based on methylation and expression, further strengthened by other positive findings - notably dysregulation in dysplasia and colorectal cancer in IBD, an established role in T cell differentiation, and protection from DSS-induced fatal colitis by MIR21 knockout. These data demonstrate a novel approach for identifying biological variations associated with germ-line variants identified by GWAS, and demonstrate translational potential for biomarker development and therapeutic target discovery.

B.2 ECCO abstract

Epigenome-wide analysis in childhood-onset Crohn's disease implicates MIR21 in pathogenesis and identifies multiple methylation-based diagnostic biomarkers [313]

N.A. Kennedy[1,2]*, A.T. Adams[1], R. Hansen[3], N.T. Ventham[1], K.R. O'Leary1[1], H. Drummond[1], C.L. Noble[1,2], E. El-Omar[3], R.K. Russell[4], D.C. Wilson[5], E.R. Nimmo[1], G.L. Hold[3], J. Satsangi[1,2].

[1]University of Edinburgh, Gastrointestinal Unit, Centre for Genomic and Experimental Medicine, Edinburgh, United Kingdom, [2]Western General Hospital, Gastrointestinal Unit, Edinburgh, United Kingdom, [3]University of Aberdeen, Gastrointestinal Research Group, Division of Applied Medicine, Aberdeen, United Kingdom, [4]Royal Hospital for Sick Children, Paediatric Gastroenterology, Glasgow, United Kingdom, [5]University of Edinburgh, Paediatric Gastroenterology and Nutrition, Child Life and Health, Edinburgh, United Kingdom

Background: DNA methylation influences transcriptional activity and marks sites of active transcription. Technological developments allow the rapid assessment of methylation state at 450,000 sites across the genome. We have performed a three-stage retrospective case-control study with epigenome-wide analysis of leucocyte DNA methylation in paediatric discovery and replication cohorts and targeted replication in adults.

Methods: Using the Illumina 450k platform we performed epigenome-wide analysis in symptomatic children presenting for diagnostic colonoscopy, half of whom were diagnosed with Crohn's Disease (CD) and half never developed any pathology. Replication was performed in children with established CD vs. symptomatic non-disease controls. Further targeted replication by pyrosequencing was performed in adults with CD vs. healthy controls.

Results: Meta-analysis of the combined paediatric datasets ($n = 66$) with the Bonferroni correction for multiple testing identified 165 CpGs with epigenome-wide significance. 138 differentially methylated regions (DMR) were identified - defined as stretches of probes where all had false discovery rate corrected p values < 0.05 and a shared direction of change in disease. Thirdly, we demonstrate an enrichment ($p < 0.0001$) of significant methylation changes in proximity to

loci implicated by genome-wide association studies (GWAS). The strongest result by each approach was MicroRNA 21, within the autophagy gene VMP1 ($p = 1.2 \times 10^{-14}$), 48kb from GWAS SNP rs1292053 - also the strongest result in the initial discovery/replication analysis. We have replicated methylation changes at this region in adults ($p = 6.6 \times 10^{-5}$, $n = 172$), and shown altered expression in blood ($p < 0.005$, $n = 66$) and inflamed vs. non-inflamed intestinal biopsies in CD ($p = 1.4 \times 10^{-6}$, $n = 99$) but not controls ($n = 73$). Alternative lines of evidence support further study of MIR21 in IBD - e.g. dysregulation in dysplasia and colorectal cancer in IBD, an established role in T cell differentiation, and protection from DSS-induced fatal colitis by MIR21 knockout. Of immediate clinical interest, linear discriminant analysis of methylation at combinations of two probes consistently predicted disease state with high accuracy (94% sensitivity, 100% specificity) in symptomatic children presenting for diagnostic colonoscopy. This aspect has the advantages of bypassing the complications of proving causation and the impact of shifting subpopulations of cells on measured methylation.

Conclusion: These data demonstrate a novel approach for identifying biological variations associated with germ-line variants identified by GWAS; and demonstrate translational potential for biomarker development and therapeutic target discovery.

B.3 UEGW abstract

Epigenetic alterations in inflammatory bowel disease — The influence of germline variation (meQTLs) on genome-wide methylation alterations.

Adams AT[1], Kalla R[1], Nimmo ER[1], IBD Character Consortium, Vatn MH[2], Jahnsen J[3], Ricanek P[3], Halfvarson J[4], Söderholm J[5], Pierik M[6], Gomollón F[7], Gut I[8,9], D'Amato M[10,11], Satsangi J[1].

[1]Institute of Genetics and Molecular Medicine, University of Edinburgh, United Kingdom. [2]University of Oslo, Institute of Clinical Medicine, EpiGen, Campus Ahus, Norway. [3]Department of Gastroenterology, Akershus University Hospital, Norway. [4]Örebro University, Department of Gastroenterology, Faculty of Medicine and Health, Örebro, Sweden. [5]Linköping University Hospital, Department of Surgery, Linköping, Sweden. [6]Maastricht University Medical Centre, Department of Gastroenterology and Hepatology, Maastricht, Netherlands. [7]HCU "Lozano Blesa", IIS Aragón, Zaragoza, Spain. [8]CNAG-CRG, Centre for Genomic Regulation, Barcelona Institute of Science and Technology, Spain. [9]Universitat Pompeu Fabra, Barcelona, Spain [10]BioCruces Health Research Institute, Bilbao, Spain. [11]Department of Biosciences and Nutrition, Karolinska Institutet, Stockholm, Sweden.

Background:

Exploring DNA methylation in Inflammatory Bowel Disease (IBD) may provide an insight into the complex gene-environment interactions in disease pathogenesis. Our study aims to characterise disease-associated methylation changes in newly diagnosed IBD and define the contribution of genetic variation, by discovery of associated quantitative trait loci (meQTL).

Methods:

Genome-wide methylation was measured in 641 DNA samples from peripheral blood (298 controls, 150 CD, 167 UC, 26 IBDU) using the Illumina 450k platform with covariates of age, sex, and cell counts, deconvoluted by the Houseman method.

Genotyping was performed using Illumina HumanOmniExpressExome8 BeadChips. Samples were obtained from new onset IBD cases in six European centres as part of the IBD Character project.

Results:

195 probes exhibited Bonferroni significant IBD-associated methylation differences, including VMP1/MIR21 ($p=3.7\times10^{-20}$), RPS6KA2 (1.1×10^{-19}), SBNO2 (2.7×10^{-19}), and TNFSF10 (1.1×10^{-15}); data which provide important replication and confirmation of methylation differences previously reported in paediatric CD[1] and adult IBD[2]. Novel findings include PHOSPHO1 (1.3×10^{-15}), MUC4 (5.5×10^{-15}), and CDH24 (1.7×10^{-14}). 1709 differentially methylated regions of consecutive FDR significant probes were defined including VMP1/MIR21, ITGB2, TNF and throughout the HLA region.

Results were highly similar in CD and UC, and only one probe showed a significant difference (NAV2, 6.82×10^{-8}). Paired genetic and methylation data showed 2327 FDR significant MeQTLs indicating a genetic influence on several key loci — RPS6KA2 (8.6×10^{-34}), ITGB2 (3.3×10^{-38}), and a replication of two SNPs previously described as correlated to VMP1/MIR21 methylation (rs8078424, $p=4.4\times10^{-25}$, rs10853015, $p=7.4\times10^{-21}$)[2].

Previously published two-probe methylation biomarkers accurately distinguished IBD from controls in this cohort (AUC $\leqslant$0.82).

Conclusions:

These data allow profiling of the IBD methylome, involving novel associations and important unequivocal replication of recent discoveries, together with insight into the genetic contribution to epigenetic alterations in complex disease.

References:

1. Adams AT, *et al.* Two-stage genome-wide methylation profiling in childhood onset Crohn's disease implicates epigenetic alterations at the VMP1/MIR21 and HLA loci. *Inflamm Bowel Dis.* 2014;20(10):1784-1793.

2. Ventham NT, *et al.* Comprehensive epigenome-wide DNA methylation profiling in inflammatory bowel disease, *Gastroenterology.* 2016;150(4):S156-7

Bibliography

[1] Abubakar I, Myhill D, Aliyu SH, *et al.* Detection of Mycobacterium avium subspecies paratuberculosis from patients with Crohn's disease using nucleic acid-based techniques: a systematic review and meta-analysis. *Inflammatory bowel diseases*, 2008;14(3):401–410.

[2] Adams A, Kennedy NA, Hansen R, *et al.* OC-003 Genome-wide Epigenetic Analysis In Childhood-onset Crohn's Disease Implicates Mir21. *Gut*, 2014;63(Suppl 1):A2.

[3] Adams A, Ventham N, Kennedy NA, *et al.* Methods and uses for determining the presence of inflammatory bowel disease, 2015.

[4] Adams AT, Kennedy NA, Hansen R, *et al.* Two-stage Genome-wide Methylation Profiling in Childhood-onset Crohn's Disease Implicates Epigenetic Alterations at the VMP1/MIR21 and HLA Loci. *Inflammatory bowel diseases*, 2014; 20(10):1784–93.

[5] Adams AT, Kennedy NA, Nimmo ER, *et al.* MEDICAMENTS, USES AND METHODS, 2014.

[6] Al Muftah WA, Al-Shafai M, Zaghlool SB, *et al.* Epigenetic associations of type 2 diabetes and BMI in an Arab population. *Clinical Epigenetics*, 2016;8(1):13.

[7] Alenghat T, Osborne LC, Saenz SA, *et al.* Histone deacetylase 3 coordinates commensal-bacteria-dependent intestinal homeostasis. *Nature*, 2013; 504(7478):153–7.

[8] Alkhouri RH, Hashmi H, Baker RD, *et al.* Vitamin and mineral status in patients with inflammatory bowel disease. *Journal of pediatric gastroenterology and nutrition*, 2013;56(1):89–92.

[9] Allchin W. Ulcerative Colitis. *Proc R Soc Med*, 1909;2:59–75.

[10] Aloi M, Tromba L, Di Nardo G, *et al.* Premature Subclinical Atherosclerosis in Pediatric Inflammatory Bowel Disease. *The Journal of Pediatrics*, 2012; 161(4):589–594.e1.

[11] Ameur A. BCRANK: Predicting binding site consensus from ranked DNA sequences, 2010.

[12] Ananthakrishnan AN, Higuchi LM, Huang ES, *et al.* Aspirin, nonsteroidal anti-inflammatory drug use, and risk for Crohn disease and ulcerative colitis: a cohort study. *Annals of internal medicine*, 2012;156(5):350–9.

[13] Ananthakrishnan AN, Khalili H, Higuchi LM, *et al.* Higher predicted vitamin D status is associated with reduced risk of Crohn's disease. *Gastroenterology*, 2012; 142(3):482–9.

[14] Ananthakrishnan AN, Khalili H, Konijeti GG, *et al.* A prospective study of long-term intake of dietary fiber and risk of Crohn's disease and ulcerative colitis. *Gastroenterology*, 2013;145(5):970–7.

[15] Ananthakrishnan AN, McGinley EL, Binion DG, *et al.* Ambient air pollution correlates with hospitalizations for inflammatory bowel disease: an ecologic analysis. *Inflammatory bowel diseases*, 2011;17(5):1138–45.

[16] Anderson CA, Boucher G, Lees CW, *et al.* Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. *Nature genetics*, 2011;43(3):246–252.

[17] Anderson CA, Massey DCO, Barrett JC, *et al.* Investigation of Crohn's Disease Risk Loci in Ulcerative Colitis Further Defines Their Molecular Relationship. *Gastroenterology*, 2009;136(2):523–529.e3.

[18] Anderson OS, Sant KE, and Dolinoy DC. Nutrition and epigenetics: an interplay of dietary methyl donors, one-carbon metabolism and DNA methylation. *The Journal of nutritional biochemistry*, 2012;23(8):853–9.

[19] Anjum R and Blenis J. The RSK family of kinases: emerging roles in cellular signalling. *Nature reviews. Molecular cell biology*, 2008;9(10):747–58.

[20] Annese V, Latiano A, Bovio P, *et al.* Genetic analysis in Italian families with inflammatory bowel disease supports linkage to the IBD1 locus–a GISC study. *Eur J Hum Genet*, 1999;7(5):567–573.

[21] Arasaradnam R, Commane D, Bradburn D, *et al.* A review of dietary factors and its influence on DNA methylation in colorectal carcinogenesis. *Epigenetics*, 2014;3(4):193–198.

[22] Arasaradnam RP, Khoo K, Bradburn M, *et al.* DNA methylation of ESR-1 and N-33 in colorectal mucosa of patients with ulcerative colitis (UC). *Epigenetics : official journal of the DNA Methylation Society*, 2010;5(5):422–6.

[23] Armitage EL, Aldhous MC, Anderson N, *et al.* Incidence of juvenile-onset Crohn's disease in Scotland: Association with northern latitude and affluence. *Gastroenterology*, 2004;127(4):1051–1057.

[24] Arnaudo AM and Garcia BA. Proteomic characterization of novel histone post-translational modifications. *Epigenetics & chromatin*, 2013;6(1):24.

[25] Aryee MJ, Jaffe AE, Corrada-Bravo H, *et al.* Minfi: A flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics*, 2014;30(10):1363–1369.

[26] Asada K, Nakajima T, Shimazu T, *et al.* Demonstration of the usefulness of epigenetic cancer risk prediction by a multicentre prospective cohort study. *Gut*, 2015;64(3):388–96.

[27] Asano K, Matsushita T, Umeno J, *et al.* A genome-wide association study identifies three new susceptibility loci for ulcerative colitis in the Japanese population. *Nature genetics*, 2009;41(12):1325–1329.

[28] Assoian RK. Common sense signalling. *Nature cell biology*, 2002;4(8):E187–E188.

[29] Atreya R, Bülte M, Gerlach GF, *et al.* Facts, myths and hypotheses on the zoonotic nature of Mycobacterium avium subspecies paratuberculosis. *International journal of medical microbiology : IJMM*, 2014;304(7):858–67.

[30] Avalos JL, Bever KM, and Wolberger C. Mechanism of sirtuin nhibition by nicotinamide: Altering the NAD + cosubstrate specificity of a Sir2 enzyme. *Molecular Cell*, 2005;17(6):855–868.

[31] Azarschab P, Porschen R, Gregor M, *et al.* Epigenetic control of the E-cadherin gene (CDH1) by CpG methylation in colectomy samples of patients with ulcerative colitis. *Genes Chromosomes and Cancer*, 2002;35(2):121–126.

[32] Azuara D, Rodriguez-Moranta F, de Oca J, *et al.* Novel methylation panel for the early detection of neoplasia in high-risk ulcerative colitis and Crohn's colitis patients. *Inflammatory bowel diseases*, 2013;19(1):165–73.

[33] Bader AG, Brown D, and Winkler M. The promise of microRNA replacement therapy. *Cancer research*, 2010;70(18):7027–30.

[34] Bae JH, Park J, Yang KM, *et al.* Detection of DNA hypermethylation in sera of patients with Crohn's disease. *Molecular medicine reports*, 2014;9(2):725–9.

[35] Barclay AR, Russell RK, Wilson ML, *et al.* Systematic Review: The Role of Breastfeeding in the Development of Pediatric Inflammatory Bowel Disease. *The Journal of Pediatrics*, 2009;155(3):421–426.

[36] Bard F, Casano L, Mallabiabarrena A, *et al.* Functional genomics reveals genes involved in protein secretion and Golgi organization. *Nature*, 2006;439(7076):604–7.

[37] Barmada MM, Brant SR, Nicolae DL, *et al.* A genome scan in 260 inflammatory bowel disease-affected relative pairs. *Inflammatory bowel diseases*, 2004; 10(5):513–520.

[38] Baron JH. Inflammatory bowel disease up to 1932. *The Mount Sinai journal of medicine, New York*, 2000;67(3):174–89.

[39] Baron S, Turck D, Leplat C, *et al.* Environmental risk factors in paediatric inflammatory bowel diseases: a population based case control study. *Gut*, 2005; 54(3):357–363.

[40] Barreiro-de Acosta M, Alvarez Castro A, Souto R, *et al.* Emigration to western industrialized countries: A risk factor for developing inflammatory bowel disease. *Journal of Crohn's & colitis*, 2011;5(6):566–9.

[41] Barrett JC, Hansoul S, Nicolae DL, *et al.* Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nature genetics*, 2008; 40(8):955–962.

[42] Bartel DP. MicroRNAs: target recognition and regulatory functions. *Cell*, 2009; 136(2):215–33.

[43] Barthel C, Spalinger MR, Brunner J, *et al.* A distinct pattern of disease-associated single nucleotide polymorphisms in IBD risk genes in a family with Crohn's disease. *European journal of gastroenterology & hepatology*, 2014; 26(7):803–6.

[44] Baylin S and Bestor TH. Altered methylation patterns in cancer cell genomes: cause or consequence? *Cancer cell*, 2002;1(4):299–305.

[45] Beamish LA, Osornio-Vargas AR, and Wine E. Air pollution: An environmental factor contributing to intestinal disease. *Journal of Crohn's & colitis*, 2011; 5(4):279–86.

[46] Beaugerie L and Sokol H. Appendicitis, not appendectomy, is protective against ulcerative colitis, both in the general population and first-degree relatives of patients with IBD. *Inflamm Bowel Dis*, 2010;16(2):356–357.

[47] Belkaid Y and Hand TW. Role of the microbiota in immunity and inflammation. *Cell*, 2014;157(1):121–41.

[48] Bellacosa A and Drohat AC. Role of base excision repair in maintaining the genetic and epigenetic integrity of CpG sites. *DNA repair*, 2015;.

[49] Benchimol EI, Fortinsky KJ, Gozdyra P, *et al.* Epidemiology of pediatric inflammatory bowel disease: a systematic review of international trends. *Inflammatory bowel diseases*, 2011;17(1):423–39.

[50] Benchimol EI, Guttmann A, Griffiths AM, *et al.* Increasing incidence of paediatric inflammatory bowel disease in Ontario, Canada: Evidence from health administrative data. *Gut*, 2009;58(11):1490–7.

[51] Benchimol EI, Mack DR, Guttmann A, *et al.* Inflammatory bowel disease in immigrants to Canada and their children: a population-based cohort study. *The American journal of gastroenterology*, 2015;110(4):553–63.

[52] Benjamini Y and Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 1995;57(1):289–300.

[53] Benko S, Magalhaes JG, Philpott DJ, *et al.* NLRC5 Limits the Activation of Inflammatory Pathways. *The Journal of Immunology*, 2010;185(3):1681–1691.

[54] Bernstein CN, Sargent M, and Rector E. Alteration in Expression of $\beta 2$ Integrins on Lamina Propria Lymphocytes in Ulcerative Colitis and Crohn's Disease. *Clinical Immunology*, 2002;104(1):67–72.

[55] Bernstein CN and Shanahan F. Disorders of a modern lifestyle: reconciling the epidemiology of inflammatory bowel diseases. *Gut*, 2008;57(9):1185–91.

[56] Bertolacini CDP, Ribeiro-Bicudo LA, Petrin A, *et al.* Clinical findings in patients with GLI2 mutations–phenotypic variability. *Clinical genetics*, 2012;81(1):70–5.

[57] Beser OF, Conde CD, Serwas NK, *et al.* Clinical features of interleukin 10 receptor gene mutations in children with very early-onset inflammatory bowel disease. *Journal of pediatric gastroenterology and nutrition*, 2015;60(3):332–8.

[58] Betel D, Koppal A, Agius P, *et al.* Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. *Genome Biology*, 2010; 11(8):R90.

[59] Bettenworth D, Nowacki TM, Ross M, *et al.* Nicotinamide treatment ameliorates the course of experimental colitis mediated by enhanced neutrophil-specific antibacterial clearance. *Molecular nutrition & food research*, 2014;58(7):1474–90.

[60] Bewtra M, Kaiser LM, TenHave T, *et al.* Crohn's disease and ulcerative colitis are associated with elevated standardized mortality ratios: a meta-analysis. *Inflammatory bowel diseases*, 2013;19(3):599–613.

[61] Bhatty M, Fan R, Muir WM, *et al.* Transcriptomic analysis of peritoneal cells in a mouse model of sepsis: confirmatory and novel results in early and late sepsis. *BMC genomics*, 2012;13(1):509.

[62] Biever A, Valjent E, and Puighermanal E. Ribosomal Protein S6 Phosphorylation in the Nervous System: From Regulation to Function. *Frontiers in Molecular Neuroscience*, 2015;8(December):1–14.

[63] Bignone PA, Lee KY, Liu Y, *et al.* RPS6KA2, a putative tumour suppressor gene at 6q27 in sporadic epithelial ovarian cancer. *Oncogene*, 2007;26(5):683–700.

[64] Binder AM, Michels KB, Rieder M, *et al.* The causal effect of red blood cell folate on genome-wide methylation in cord blood: a Mendelian randomization approach. *BMC Bioinformatics*, 2013;14(1):353.

[65] Bird A. CpG-rich islands and the function of DNA methylation. *Nature*, 1986; 321(6067):209–13.

[66] Bird A. Perceptions of epigenetics. *Nature*, 2007;447(7143):396–8.

[67] Bis JC, DeCarli C, Smith AV, *et al.* Common variants at 12q14 and 12q24 are associated with infant head circumference. *Nature genetics*, 2012;44(5):545–51.

[68] Bode JG, Nimmesgern A, Schmitz J, *et al.* LPS and TNFα induce SOCS3 mRNA and inhibit IL-6-induced activation of STAT3 in macrophages. *FEBS Letters*, 1999;463(3):365–370.

[69] Bone KR, Gruper Y, Goldenberg D, *et al.* Translation regulation of Runx3. *Blood cells, molecules & diseases*, 2010;45(2):112–6.

[70] Boonstra K, de Vries EM, van Geloven N, *et al.* Risk factors for Primary Sclerosing Cholangitis. *Liver International*, 2015;pp. 1–8.

[71] Boosani CS and Agrawal DK. Methylation and microRNA-mediated epigenetic regulation of SOCS3. *Molecular biology reports*, 2015;42(4):853–872.

[72] Borba EF, Carvalho JF, and Bonfá E. Mechanisms of Dyslipoproteinemias in Systemic Lupus Erythematosus. *Clinical and Developmental Immunology*, 2006; 13(2-4):203–208.

[73] Boyle JP, Parkhouse R, and Monie TP. Insights into the molecular basis of the NOD2 signalling pathway. *Open biology*, 2014;4(12):140178–.

[74] Brain O, Owens BMJ, Pichulik T, *et al.* The Intracellular Sensor NOD2 Induces MicroRNA-29 Expression in Human Dendritic Cells to Limit IL-23 Release. *Immunity*, 2013;39(3):521–36.

[75] Brant SR, Fu Y, Fields CT, *et al.* American families with Crohn's disease have strong evidence for linkage to chromosome 16 but not chromosome 12. *Gastroenterology*, 1998;115:1056–1061.

[76] Brenner O, Levanon D, Negreanu V, *et al.* Loss of Runx3 function in leukocytes is associated with spontaneously developed colitis and gastric mucosal hyperplasia. *Proceedings of the National Academy of Sciences of the United States of America*, 2004;101(45):16016–16021.

[77] Brinar M, Čuković-Čavka S, Bozina N, *et al.* MDR1 polymorphisms are associated with inflammatory bowel disease in a cohort of Croatian IBD patients. *BMC gastroenterology*, 2013;13(1):57.

[78] Buck AH, Coakley G, Simbari F, *et al.* Exosomes secreted by nematode parasites transfer small RNAs to mammalian cells and modulate innate immunity. *Nat Commun*, 2014;5:5488.

[79] Buck-Koehntop BA and Defossez PA. On how mammalian transcription factors recognize methylated DNA. *Epigenetics : official journal of the DNA Methylation Society*, 2013;8(2):131–7.

[80] Burger L, Gaidatzis D, Schubeler D, *et al.* Identification of active regulatory regions from DNA methylation data. *Nucleic Acids Research*, 2013;41(16):e155.

[81] Burgess RJ and Zhang Z. Histone chaperones in nucleosome assembly and human disease. *Nature structural & molecular biology*, 2013;20(1):14–22.

[82] Burisch J and Munkholm P. Inflammatory bowel disease epidemiology. *Current opinion in gastroenterology*, 2013;29(4):357–62.

[83] Burisch J, Pedersen N, Čuković-Čavka S, *et al.* East-West gradient in the incidence of inflammatory bowel disease in Europe: the ECCO-EpiCom inception cohort. *Gut*, 2014;63(4):588–97.

[84] Burisch J, Pedersen N, Čuković-Čavka S, *et al.* Environmental factors in a population-based inception cohort of inflammatory bowel disease patients in Europe - An ECCO-EpiCom study. *Journal of Crohn's and Colitis*, 2014;8(7):607–616.

[85] Buruiana FE, Solà I, and Alonso-Coello P. Recombinant human interleukin 10 for induction of remission in Crohn's disease. *Cochrane database of systematic reviews (Online)*, 2010;(11):CD005109.

[86] Cadwell K, Liu JY, Brown SL, *et al.* A key role for autophagy and the autophagy gene Atg16l1 in mouse and human intestinal Paneth cells. *Nature*, 2008; 456(7219):259–63.

[87] Cai X, Hagedorn CH, and Cullen BR. Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *RNA*, 2004; 10(12):1957–1966.

[88] Calkins BM. A meta-analysis of the role of smoking in inflammatory bowel disease. *Digestive diseases and sciences*, 1989;34(12):1841–54.

[89] Calvo-Garrido J, Carilla-Latorre S, Lázaro-Diéguez F, *et al.* Vacuole membrane protein 1 is an endoplasmic reticulum protein required for organelle biogenesis, protein secretion, and development. *Molecular biology of the cell*, 2008; 19(8):3442–53.

[90] Calvo-Garrido J and Escalante R. Autophagy dysfunction and ubiquitin-positive protein aggregates in Dictyostelium cells lacking Vmp1. *Autophagy*, 2010; 6(1):100–9.

[91] Calvo-Garrido J, King JS, Muñoz-Braceras S, *et al.* Vmp1 Regulates PtdIns3P Signaling During Autophagosome Formation in Dictyostelium discoideum. *Traffic*, 2014;15(11):1235–1246.

[92] Cameron EE, Bachman KE, Myöhänen S, *et al.* Synergy of demethylation and histone deacetylase inhibition in the re-expression of genes silenced in cancer. *Nature genetics*, 1999;21(1):103–7.

[93] Cannioto Z, Berti I, Martelossi S, *et al.* IBD and IBD mimicking enterocolitis in children younger than 2 years of age. *European Journal of Pediatrics*, 2009; 168:149–155.

[94] Capelluto DGS. Tollip: a multitasking protein in innate immunity and protein trafficking. *Microbes and infection / Institut Pasteur*, 2012;14(2):140–7.

[95] Cappellano G, Orilieri E, Comi C, *et al.* Variations of the perforin gene in patients with multiple sclerosis. *Genes and immunity*, 2008;9(April):438–444.

[96] Carey V. ROC: utilities for ROC with uarray focus.

[97] Carey V. gwascat, 2015.

[98] Carletti MZ, Fiedler SD, and Christenson LK. MicroRNA 21 blocks apoptosis in mouse periovulatory granulosa cells. *Biology of reproduction*, 2010;83(2):286–95.

[99] Carmona FJ, Azuara D, Berenguer-Llergo A, *et al.* DNA methylation biomarkers for noninvasive diagnosis of colorectal cancer. *Cancer prevention research*, 2013; 6(7):656–65.

[100] Carson HJ, Dudley MH, Knight LD, *et al.* Psychosocial Complications of Crohn's Disease and Cause of Death. *Journal of Forensic Sciences*, 2014;59(2):568–570.

[101] Casanova JL and Abel L. Revisiting Crohn's disease as a primary immunodeficiency of macrophages. *Journal of Experimental Medicine*, 2009;206(9):1839–1843.

[102] Casella G, Tontini GE, Bassotti G, *et al.* Neurological disorders and inflammatory bowel diseases. *World Journal of Gastroenterology*, 2014;20(27):8764–8782.

[103] Celiberto LS, Bedani R, Rossi EA, *et al.* Probiotics: The Scientific Evidence in the Context of Inflammatory Bowel Disease. *Critical Reviews in Food Science and Nutrition*, 2015;8398(November).

[104] Chang CC, Zhang QY, Liu Z, *et al.* Downregulation of inflammatory microRNAs by Ig-like transcript 3 is essential for the differentiation of human CD8(+) T suppressor cells. *Journal of immunology (Baltimore, Md. : 1950)*, 2012;188(7):3042–52.

[105] Chang PV, Hao L, Offermanns S, *et al.* The microbial metabolite butyrate regulates intestinal macrophage function via histone deacetylase inhibition. *Proceedings of the National Academy of Sciences of the United States of America*, 2014; 111(6):2247–52.

[106] Chaudrey K, Salvaggio M, Ahmed A, *et al.* Updates in vaccination: recommendations for adult inflammatory bowel disease patients. *World journal of gastroenterology : WJG*, 2015;21(11):3184–96.

[107] Chauhan S, Mandell MA, and Deretic V. IRGM Governs the Core Autophagy Machinery to Conduct Antimicrobial Defense. *Molecular cell*, 2015;58(3):507–521.

[108] Chen M, Peyrin-Biroulet L, George A, *et al.* Methyl deficient diet aggravates experimental colitis in rats. *Journal of cellular and molecular medicine*, 2011; 15(11):2486–97.

[109] Chen RH, Sarnecki C, and Blenis J. Nuclear localization and regulation of erk- and rsk-encoded protein kinases. *Molecular and cellular biology*, 1992;12(3):915–927.

[110] Chen X, Zen K, and Zhang CY. Reply to Lack of detectable oral bioavailability of plant microRNAs after feeding in mice. *Nature biotechnology*, 2013;31(11):967–9.

[111] Chen Y, Wang C, Liu Y, *et al.* miR-122 targets NOD2 to decrease intestinal epithelial cell injury in Crohn's disease. *Biochemical and biophysical research communications*, 2013;438(1):133–9.

[112] Cheng X, Zhang X, Su J, *et al.* miR-19b downregulates intestinal SOCS3 to reduce intestinal inflammation in Crohn's disease. *Scientific Reports*, 2015;5:10397.

[113] Chiba T, Marusawa H, and Ushijima T. Inflammation-associated cancer development in digestive organs: Mechanisms and roles for genetic and epigenetic modulation. *Gastroenterology*, 2012;143(3):550–563.

[114] Chiodini R. Possible Role of Mycobacteria in Inflammatory Bowel-Disease .1. an Unclassified Mycobacterium Species Isolated From Patients With Crohns-Disease. *Digestive diseases and sciences*, 1984;29(12):1073–1079.

[115] Cho JH, Nicolae DL, Gold LH, *et al.* Identification of novel susceptibility loci for inflammatory bowel disease on chromosomes 1p, 3q, and 4q: evidence for epistasis between 1p and IBD1. *Proceedings of the National Academy of Sciences of the United States of America*, 1998;95(13):7502–7.

[116] Cho JH, Nicolae DL, Ramos R, *et al.* Linkage and linkage disequilibrium in chromosome band 1p36 in American Chaldeans with inflammatory bowel disease. *Human molecular genetics*, 2000;9(9):1425–32.

[117] Chuang AY, Chuang JC, Zhai Z, *et al.* NOD2 Expression is Regulated by microRNAs in Colonic Epithelial HCT116 Cells. *Inflammatory bowel diseases*, 2014; 20(1):126–35.

[118] Chung DD, Honda K, Cafuir L, *et al.* The Runx3 distal transcript encodes an additional transcriptional activation domain. *The FEBS journal*, 2007; 274(13):3429–39.

[119] Church TR, Wandell M, Lofton-Day C, *et al.* Prospective evaluation of methylated SEPT9 in plasma for detection of asymptomatic colorectal cancer. *Gut*, 2014;63(2):317–25.

[120] Clemente JC, Ursell LK, Parfrey LW, *et al.* The impact of the gut microbiota on human health: an integrative view. *Cell*, 2012;148(6):1258–70.

[121] Cohen LE. GLI2 mutations as a cause of hypopituitarism. *Pediatric endocrinology reviews : PER*, 2012;9(4):706–9.

[122] Cohney SJ, Sanden D, Cacalano NA, *et al.* SOCS-3 is tyrosine phosphorylated in response to interleukin-2 and suppresses STAT5 phosphorylation and lymphocyte proliferation. *Molecular and cellular biology*, 1999;19(7):4980–4988.

[123] Colman RJ and Rubin DT. Fecal microbiota transplantation as therapy for inflammatory bowel disease: a systematic review and meta-analysis. *Journal of Crohn's & colitis*, 2014;8(12):1569–81.

[124] Connelly TM, Berg AS, Harris L, *et al.* Genetic Determinants Associated With Early Age of Diagnosis of IBD. *Diseases of the Colon & Rectum*, 2015;58:321–327.

[125] Cooke J, Zhang H, Greger L, *et al.* Mucosal genome-wide methylation changes in inflammatory bowel disease. *Inflammatory bowel diseases*, 2012;18(11):2128–37.

[126] Cornish JA, Tan E, Simillis C, *et al.* The risk of oral contraceptives in the etiology of inflammatory bowel disease: a meta-analysis. *The American journal of gastroenterology*, 2008;103(9):2394–400.

[127] Correa I, Veny M, Esteller M, *et al.* Defective IL-10 production in severe phenotypes of Crohn's disease. *Journal of Leukocyte Biology*, 2009;85(5):896–903.

[128] Coskun M, Bjerrum JT, Seidelin JB, *et al.* miR-20b, miR-98, miR-125b-1*, and let-7e* as new potential diagnostic biomarkers in ulcerative colitis. *World Journal of Gastroenterology*, 2013;19(27):4289–4299.

[129] Cosnes J, Carbonnel F, Beaugerie L, *et al.* Effects of appendicectomy on the course of ulcerative colitis. *Gut*, 2002;51(6):803–807.

[130] Craig G. Alfred the Great: a diagnosis. *Journal of the Royal Society of Medicine*, 1991;84(5):303–305.

[131] Cribari-Neto F and Zeileis A. Beta regression in R. *Journal Of Statistical Software*, 2010;34(2).

[132] Crockett SD, Hansen RA, Stürmer T, *et al.* Statins are associated with reduced use of steroids in inflammatory bowel disease: A retrospective cohort study. *Inflammatory Bowel Diseases*, 2012;18(6):1048–1056.

[133] Crohn BB. An historic note on ulcerative colitis. *Gastroenterology*, 1962;42:366.

[134] Crohn BB, Ginzburg L, and Oppenheimer GD. Regional Ileitis: A Pathologic and Clinical Entity. *Journal of the American Medical Association*, 1932;99:1323–1329.

[135] Croker BA, Krebs DL, Zhang JG, *et al.* SOCS3 negatively regulates IL-6 signaling in vivo. *Nature immunology*, 2003;4(6):540–545.

[136] Cui J, Zhu L, Xia X, *et al.* NLRC5 negatively regulates the NF-kappaB and type I interferon signaling pathways. *Cell*, 2010;141(3):483–96.

[137] Cummings JRF, Cooney RM, Clarke G, *et al.* The genetics of NOD-like receptors in Crohn's disease. *Tissue Antigens*, 2010;18(7):48–56.

[138] Curran ME, Lau KF, Hampe J, *et al.* Genetic analysis of inflammatory bowel disease in a large European cohort supports linkage to chromosomes 12 and 16. *Gastroenterology*, 1998;115(5):1066–1071.

[139] Dalby KN, Morrice N, Barry F, *et al.* Identification of Regulatory Phosphorylation Sites in Mitogen- activated Protein Kinase (MAPK) -activated Protein Kinase-1a / p90 rsk That Are Inducible by MAPK *. *The Journal of biological chemistry*, 1998;273(3):1496–1505.

[140] Dalmasso G, Nguyen HTT, Yan Y, *et al.* Microbiota modulate host gene expression via microRNAs. *PloS one*, 2011;6(4):e19293.

[141] Dalziel T. Chronic Interstitial Enteritis. *British Medical Journal*, 1913;2:1068–1070.

[142] Danoy P, Pryce K, Hadler J, *et al.* Association of variants at 1q32 and STAT3 with ankylosing spondylitis suggests genetic overlap with Crohn's disease. *PLoS Genetics*, 2010;6(12):1–5.

[143] Darfeuille-Michaud A, Boudeau J, Bulois P, *et al.* High prevalence of adherent-invasive Escherichia coli associated with ileal mucosa in Crohn's disease. *Gastroenterology*, 2004;127(2):412–421.

[144] De Preter V, Arijs I, Windey K, *et al.* Impaired butyrate oxidation in ulcerative colitis is due to decreased butyrate uptake and a defect in the oxidation pathway. *Inflammatory bowel diseases*, 2012;18(6):1127–36.

[145] De Preter V, Rutgeerts P, Schuit F, *et al.* Impaired expression of genes involved in the butyrate oxidation pathway in Crohn's disease patients. *Inflammatory bowel diseases*, 2013;19(3):E43–4.

[146] Denizot J, Desrichard A, Agus A, *et al.* Diet-induced hypoxia responsive element demethylation increases CEACAM6 expression, favouring Crohn's disease-associated Escherichia coli colonisation. *Gut*, 2015;64(3):428–37.

[147] Dennler S, André J, Alexaki I, *et al.* Induction of sonic hedgehog mediators by transforming growth factor-β: Smad3-dependent activation of Gli2 and Gli1 expression in vitro and in vivo. *Cancer Research*, 2007;67(14):6981–6986.

[148] Devaraj B and Kaiser AM. Surgical Management of Ulcerative Colitis in the Era of Biologicals. *Inflammatory Bowel Diseases*, 2015;21(1):208–220.

[149] Dey BR, Furlanetto RW, and Nissley P. Suppressor of cytokine signaling (SOCS)-3 protein interacts with the insulin-like growth factor-I receptor. *Biochemical and biophysical research communications*, 2000;278(1):38–43.

[150] Dianda L, Hanby AM, Wright NA, *et al.* T cell receptor-alpha beta-deficient mice fail to develop colitis in the absence of a microbial environment. *The American journal of pathology*, 1997;150(1):91–97.

[151] Dickinson B, Zhang Y, Petrick JS, *et al.* Lack of detectable oral bioavailability of plant microRNAs after feeding in mice. *Nature biotechnology*, 2013;31(11):965–7.

[152] Didonato JA, Mercurio F, and Karin M. NF-kB and the link between inflammation and cancer. *Immunological Reviews*, 2012;246(1):379–400.

[153] Djebali S, Davis CA, Merkel A, *et al.* Landscape of transcription in human cells. *Nature*, 2012;489(7414):101–8.

[154] Djuretic IM, Levanon D, Negreanu V, *et al.* Transcription factors T-bet and Runx3 cooperate to activate Ifng and silence Il4 in T helper type 1 cells. *Nature immunology*, 2007;8(2):145–153.

[155] Dombal DE. Ulcerative colitis : definition , historical background , aetiology , diagnosis , natural history and local complications. *Postgraduate Medical Journal*, 1968;44(515):684–692.

[156] Du P, Kibbe WA, and Lin SM. lumi: A pipeline for processing Illumina microarray. *Bioinformatics*, 2008;24(13):1547–1548.

[157] Du P, Zhang X, Huang CC, *et al.* Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC bioinformatics*, 2010;11(1):587.

[158] Duerr RH, Taylor KD, Brant SR, *et al.* A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. *Science (New York, N.Y.)*, 2006; 314(5804):1461–1463.

[159] Duggan AE, Usmani I, Neal KR, *et al.* Appendicectomy, childhood hygiene, Helicobacter pylori status, and risk of inflammatory bowel disease: a case control study. *Gut*, 1998;43(4):494–8.

[160] Dümmler BA, Hauge C, Silber J, *et al.* Functional characterization of human RSK4, a new 90-kDa ribosomal S6 kinase, reveals constitutive activation in most cell types. *Journal of Biological Chemistry*, 2005;280(14):13304–13314.

[161] Dunkley TPJ, Hester S, Shadforth IP, *et al.* Mapping the Arabidopsis organelle proteome. *Proceedings of the National Academy of Sciences of the United States of America*, 2006;103(17):6518–23.

[162] Durães C, Machado JC, Portela F, *et al.* Phenotype-genotype profiles in Crohn's disease predicted by genetic markers in autophagy-related genes (GOIA Study II). *Inflammatory Bowel Diseases*, 2013;19(2):230–239.

[163] Dusetti NJ, Jiang Y, Vaccaro MI, *et al.* Cloning and expression of the rat vacuole membrane protein 1 (VMP1), a new gene activated in pancreas with acute pancreatitis, which promotes vacuole formation. *Biochemical and biophysical research communications*, 2002;290(2):641–9.

[164] Eads CA, Lord RV, Kurumboor SK, *et al.* Fields of Aberrant CpG Island Hypermethylation in Barrett's Esophagus and Associated Adenocarcinoma. *Cancer research*, 2000;60(18):5021–5026.

[165] Ehrlich M. DNA hypomethylation in cancer cells. *Epigenomics*, 2009;1(2):239–259.

[166] Eichler EE, Flint J, Gibson G, *et al.* Missing heritability and strategies for finding the underlying causes of complex disease. *Nature reviews. Genetics*, 2010;11(6):446–50.

[167] Eidson M, Wahlstrom J, Beaulieu AM, *et al.* Altered development of NKT cells, $\gamma\delta$ T cells, CD8 T cells and NK cells in a PLZF deficient patient. *PloS one*, 2011;6(9):e24441.

[168] Ek WE, D'Amato M, and Halfvarson J. The history of genetics in inflammatory bowel disease. *Annals of gastroenterology*, 2014;27(4):294–303.

[169] Ekbom A, Helmick C, Zack M, *et al.* The epidemiology of inflammatory bowel disease: a large, population-based study in Sweden. *Gastroenterology*, 1991;100(2):350–8.

[170] El Kasmi KC, Holst J, Coffre M, *et al.* General Nature of the STAT3-Activated Anti-Inflammatory Response. *The Journal of Immunology*, 2006;177(11):7880–7888.

[171] El Kasmi KC, Smith AM, Williams L, *et al.* Cutting Edge: A Transcriptional Repressor and Corepressor Induced by the STAT3-Regulated Anti-Inflammatory Signaling Pathway. *The Journal of Immunology*, 2007;179(11):7215–7219.

[172] Elenberg Y, Shani-adir A, Hecht Y, *et al.* Pyoderma gangrenosum after Bone marrow transplantation for leukocyte adhesion deficiency type 1. *Israel Medical Association Journal*, 2010;12(February):11–12.

[173] Ender C, Krek A, Friedländer MR, *et al.* A human snoRNA with microRNA-like functions. *Molecular cell*, 2008;32(4):519–28.

[174] Fabbri M, Garzon R, Cimmino A, *et al.* MicroRNA-29 family reverts aberrant methylation in lung cancer by targeting DNA methyltransferases 3A and 3B. *Proceedings of the National Academy of Sciences of the United States of America*, 2007;104(40):15805–10.

[175] Fang Y, Ren X, and Feng Z. Genetic correlation of SOCS3 polymorphisms with infantile asthma: an evidence based on a case-control study. *International journal of clinical and experimental pathology*, 2015;8(8):9586–91.

[176] Farh KK, Marson A, Zhu J, *et al.* Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature*, 2015;518(7539):337–43.

[177] Farrington P, Miller E, Calman K, *et al.* Letters To the Editor: Measles vaccination as a risk factor for inflammatory bowel disease. *Lancet*, 1995;345(8961):1362–4.

[178] Fedorak RN, Gangl A, Elson CO, *et al.* Recombinant human interleukin 10 in the treatment of patients with mild to moderately active Crohn's disease. The Interleukin 10 Inflammatory Bowel Disease Cooperative Study Group. *Gastroenterology*, 2000;119(6):1473–1482.

[179] Feldmann A, Ivanek R, Murr R, *et al.* Transcription Factor Occupancy Can Mediate Active Turnover of DNA Methylation at Regulatory Regions. *PLoS Genetics*, 2013;9(12):e1003994.

[180] Felice C, Lewis A, Armuzzi A, *et al.* Review article: selective histone deacetylase isoforms as potential therapeutic targets in inflammatory bowel diseases. *Alimentary pharmacology & therapeutics*, 2015;41(1):26–38.

[181] Feller M, Huwiler K, Stephan R, *et al.* Mycobacterium avium subspecies paratuberculosis and Crohn's disease: a systematic review and meta-analysis. *The Lancet infectious diseases*, 2007;7(9):607–613.

[182] Feng Y, He D, Yao Z, *et al.* The machinery of macroautophagy. *Cell Research*, 2014;24(1):24–41.

[183] Ferguson A and Sedgwick DM. Juvenile-onset inflammatory bowel disease: predictors of morbidity and health status in early adult life. *J R Coll Physicians Lond*, 1994;28(3):220–227.

[184] Ferrari SLP and Cribari-Neto F. Beta Regression for Modelling Rates and Proportions. *Journal of Applied Statistics*, 2004;31(7):799–815.

[185] Ferree A, Guillily M, Li H, *et al.* Regulation of physiologic actions of LRRK2: Focus on autophagy. *Neurodegenerative Diseases*, 2012;10(1-4):238–241.

[186] Firouzi F, Bahari A, Aghazadeh R, *et al.* Appendectomy, tonsillectomy, and risk of inflammatory bowel disease: A case control study in Iran. *International Journal of Colorectal Disease*, 2006;21(2):155–159.

[187] Foulks JM, Parnell KM, Nix RN, *et al.* Epigenetic drug discovery: targeting DNA methyltransferases. *Journal of biomolecular screening*, 2012;17(1):2–17.

[188] Fowler EV, Doecke J, Simms LA, *et al.* ATG16L1 T300A shows strong associations with disease subgroups in a large Australian IBD population: Further support for significant disease heterogeneity. *American Journal of Gastroenterology*, 2008;103(10):2519–2526.

[189] Fraczek J, Vanhaecke T, and Rogiers V. Toxicological and metabolic considerations for histone deacetylase inhibitors. *Expert opinion on drug metabolism & toxicology*, 2013;9(4):441–57.

[190] Frank DN, St Amand AL, Feldman RA, *et al.* Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proceedings of the National Academy of Sciences of the United States of America*, 2007;104(34):13780–13785.

[191] Franke A, Balschun T, Karlsen TH, *et al.* Sequence variants in IL10, ARPC2 and multiple other loci contribute to ulcerative colitis susceptibility. *Nature genetics*, 2008;40(11):1319–1323.

[192] Franke A, Balschun T, Sina C, *et al.* Genome-wide association study for ulcerative colitis identifies risk loci at 7q22 and 22q13 (IL17REL). *Nature genetics*, 2010; 42(4):292–294.

[193] Franke A, Hampe J, Rosenstiel P, *et al.* Systematic association mapping identifies NELL1 as a novel IBD disease gene. *PLoS ONE*, 2007;2(8).

[194] Franke A, McGovern DPB, Barrett JC, *et al.* Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nature genetics*, 2010;42(12):1118–25.

[195] Frisch M and Biggar RJ. Appendectomy and protection against ulcerative colitis. *The New England journal of medicine*, 2001;345(3):222–3.

[196] Frisch M, Pedersen BV, and Andersson RE. Appendicitis, mesenteric lymphadenitis, and subsequent risk of ulcerative colitis: cohort studies in Sweden and Denmark. *BMJ (Clinical research ed.)*, 2009;338:b716.

[197] Frobøse H, Rønn SG, Heding PE, *et al.* Suppressor of cytokine Signaling-3 inhibits interleukin-1 signaling by targeting the TRAF-6/TAK1 complex. *Molecular endocrinology (Baltimore, Md.)*, 2006;20(March):1587–1596.

[198] Fujii S, Tominaga K, Kitajima K, *et al.* Methylation of the oestrogen receptor gene in non-neoplastic epithelium as a marker of colorectal neoplasia risk in longstanding and extensive ulcerative colitis. *Gut*, 2005;54(9):1287–1292.

[199] Fujita S, Ito T, Mizutani T, *et al.* miR-21 Gene expression triggered by AP-1 is sustained through a double-negative feedback mechanism. *Journal of molecular biology*, 2008;378(3):492–504.

[200] Furusawa Y, Obata Y, Fukuda S, *et al.* Commensal microbe-derived butyrate induces the differentiation of colonic regulatory T cells. *Nature*, 2013; 504(7480):446–50.

[201] Gao L, Cueto MA, Asselbergs F, *et al.* Cloning and functional characterization of HDAC11, a novel member of the human histone deacetylase family. *The Journal of biological chemistry*, 2002;277(28):25748–55.

[202] Gardet A, Benita Y, Li C, *et al.* LRRK2 is involved in the IFN-gamma response and host response to pathogens. *Journal of immunology (Baltimore, Md. : 1950)*, 2010;185(9):5577–85.

[203] Garrity-Park MM, Loftus EV, Sandborn WJ, *et al.* Methylation status of genes in non-neoplastic mucosa from patients with ulcerative colitis-associated colorectal cancer. *The American journal of gastroenterology*, 2010;105(7):1610–9.

[204] Garzon R, Heaphy CEA, Havelange V, *et al.* MicroRNA 29b functions in acute myeloid leukemia. *Blood*, 2009;114(26):5331–41.

[205] Garzon R, Marcucci G, and Croce CM. Targeting microRNAs in cancer: rationale, strategies and challenges. *Nature reviews. Drug discovery*, 2010;9(10):775–89.

[206] Gearry RB, Richardson AK, Frampton CM, *et al.* Population-based cases control study of inflammatory bowel disease risk factors. *J Gastroenterol Hepatol*, 2010; 25(2):325–33.

[207] Geeleher P, Hartnett L, Egan LJ, *et al.* Gene-set analysis is severely biased when applied to genome-wide methylation data. *Bioinformatics (Oxford, England)*, 2013;29(15):1851–7.

[208] Gerlach UA, Vrakas G, Reddy S, *et al.* Chronic Intestinal Failure After Crohn Disease. *JAMA Surgery*, 2014;149(10):1060.

[209] Gevers D, Kugathasan S, Denson L, *et al.* The Treatment-Naive Microbiome in New-Onset Crohn's Disease. *Cell Host & Microbe*, 2014;15(3):382–392.

[210] Ghorpade DS, Sinha AY, Holla S, *et al.* NOD2-nitric oxide-responsive microRNA-146a activates Sonic hedgehog signaling to orchestrate inflammatory responses in murine model of inflammatory bowel disease. *The Journal of biological chemistry*, 2013;288(46):33037–48.

[211] Glauben R, Batra A, Fedke I, *et al.* Histone hyperacetylation is associated with amelioration of experimental colitis in mice. *Journal of immunology*, 2006; 176(8):5015–22.

[212] Glauben R, Batra A, Stroh T, *et al.* Histone deacetylases: novel targets for prevention of colitis-associated cancer in mice. *Gut*, 2008;57(5):613–22.

[213] Glocker EO, Kotlarz D, Boztug K, *et al.* Inflammatory bowel disease and mutations affecting the interleukin-10 receptor. *N Engl J Med*, 2009;361(21):2033–2045.

[214] Glória L, Cravo M, Pinto A, *et al.* DNA hypomethylation and proliferative activity are increased in the rectal mucosa of patients with long-standing ulcerative colitis. *Cancer*, 1996;78(11):2300–2306.

[215] Godnic I, Zorc M, Jevsinek Skok D, *et al.* Genome-wide and species-wide in silico screening for intragenic MicroRNAs in human, mouse and chicken. *PloS one*, 2013;8(6):e65165.

[216] Gonda TA, Kim YI, Salas MC, *et al.* Folic acid increases global DNA methylation and reduces inflammation to prevent Helicobacter-associated gastric cancer in mice. *Gastroenterology*, 2012;142(4):824–833.e7.

[217] Gonsky R, Deem RL, Landers CJ, *et al.* Distinct IFNG methylation in a subset of ulcerative colitis patients based on reactivity to microbial antigens. *Inflammatory bowel diseases*, 2011;17(1):171–8.

[218] Gonsky R, Deem RL, Landers CJ, *et al.* IFNG rs1861494 polymorphism is associated with IBD disease severity and functional changes in both IFNG methylation and protein secretion. *Inflammatory bowel diseases*, 2014;20(10):1794–801.

[219] Gonsky R, Deem RL, and Targan SR. Distinct Methylation of IFNG in the Gut. *Journal of interferon & cytokine research*, 2009;29(7):407–14.

[220] Good PJ, Guyer MS, Kamholz S, *et al.* The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science*, 2004;306(5696):636–40.

[221] Grasso D, Ropolo A, Lo Ré A, *et al.* Zymophagy, a novel selective autophagy pathway mediated by VMP1-USP9x-p62, prevents pancreatic cell death. *The Journal of biological chemistry*, 2011;286(10):8308–24.

[222] Graziewicz MA, Tarrant TK, Buckley B, *et al.* An Endogenous TNF-α Antagonist Induced by Splice-switching Oligonucleotides Reduces Inflammation in Hepatitis and Arthritis Mouse Models. *Molecular Therapy*, 2008;16(7):1316–1322.

[223] Griffiths AM. Specificities of inflammatory bowel disease in childhood. *Best practice & research. Clinical gastroenterology*, 2004;18(3):509–23.

[224] Griffiths AM. Growth retardation in early-onset inflammatory bowel disease: Should we monitor and treat these patients differently? *Digestive Diseases*, 2009;27(3):404–411.

[225] Grill M, Syme TE, Noçon AL, *et al.* Strawberry notch homolog 2 is a novel inflammatory response factor predominantly but not exclusively expressed by astrocytes in the central nervous system. *Glia*, 2015;63(10):1738–52.

[226] Gros C, Fleury L, Nahoum V, *et al.* New Insights on the Mechanism of Quinoline-based DNA Methyltransferase Inhibitors. *The Journal of biological chemistry*, 2015;290(10):6293–302.

[227] Grützmann R, Molnar B, Pilarsky C, *et al.* Sensitive detection of colorectal cancer in peripheral blood by septin 9 DNA methylation assay. *PloS one*, 2008; 3(11):e3759.

[228] Guagnozzi D and Lucendo AJ. Anemia in inflammatory bowel disease: a neglected issue with relevant effects. *World journal of gastroenterology*, 2014; 20(13):3542–51.

[229] Guo C, Ahmad T, Beckly J, *et al.* Association of caspase-9 and RUNX3 with inflammatory bowel disease. *Tissue antigens*, 2011;77(1):23–9.

[230] Guo C, Ding J, Yao L, *et al.* Tumor suppressor gene Runx3 sensitizes gastric cancer cells to chemotherapeutic drugs by downregulating Bcl-2, MDR-1 and MRP-1. *International Journal of Cancer*, 2005;116(1):155–160.

[231] Guo C, Yao F, Wu K, *et al.* Chromatin immunoprecipitation and association study revealed a possible role of Runt-related transcription factor 3 in the ulcerative colitis of Chinese population. *Clinical immunology (Orlando, Fla.)*, 2010; 135(3):483–9.

[232] Guo XZ, Ye XL, Xiao WZ, *et al.* Downregulation of VMP1 confers aggressive properties to colorectal cancer. *Oncology reports*, 2015;34(5):2557–66.

[233] Gutte PGM, Jurt S, Grütter MG, *et al.* Unusual structural features revealed by the solution NMR structure of the NLRC5 caspase recruitment domain. *Biochemistry*, 2014;53(19):3106–3117.

[234] Haider BA, Baras AS, McCall MN, *et al.* A critical evaluation of microRNA biomarkers in non-neoplastic disease. *PloS one*, 2014;9(2):e89565.

[235] Halfvarson J. Genetics in twins with Crohn's disease: Less pronounced than previously believed? *Inflammatory Bowel Diseases*, 2011;17(1):6–12.

[236] Hamer HM, Jonkers D, Venema K, *et al.* Review article: the role of butyrate on colonic function. *Alimentary pharmacology & therapeutics*, 2008;27(2):104–19.

[237] Hamer HM, Jonkers DMAE, Vanhoutvin SALW, *et al.* Effect of butyrate enemas on inflammation and antioxidant status in the colonic mucosa of patients with ulcerative colitis in remission. *Clinical nutrition*, 2010;29(6):738–44.

[238] Hampe J, Cuthbert A, Croucher PJ, *et al.* Association between insertion mutation in NOD2 gene and Crohn's disease in German and British populations. *Lancet*, 2001;357(9272):1925–8.

[239] Hampe J, Franke A, Rosenstiel P, *et al.* A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. *Nature genetics*, 2007;39(2):207–11.

[240] Hampe J, Schreiber S, Shaw SH, *et al.* A genomewide analysis provides evidence for novel linkages in inflammatory bowel disease in a large European cohort. *American journal of human genetics*, 1999;64(3):808–16.

[241] Han L, Witmer PD, Casey E, *et al.* DNA methylation regulates MicroRNA expression. *Cancer biology & therapy*, 2007;6(8):1284–8.

[242] Hanna S and Etzioni A. Leukocyte adhesion deficiencies. *Annals of the New York Academy of Sciences*, 2012;1250(1):50–55.

[243] Hansen R, Berry SH, Mukhopadhya I, *et al.* The microaerophilic microbiota of de-novo paediatric inflammatory bowel disease: the BISCUIT study. *PloS one*, 2013;8(3):e58825.

[244] Hansen R, Russell RK, Reiff C, *et al.* Microbiota of de-novo pediatric IBD: increased Faecalibacterium prausnitzii and reduced bacterial diversity in Crohn's but not in ulcerative colitis. *The American journal of gastroenterology*, 2012; 107(12):1913–22.

[245] Hansen TB, Jensen TI, Clausen BH, *et al.* Natural RNA circles function as efficient microRNA sponges. *Nature*, 2013;495(7441):384–8.

[246] Hansen TS, Jess T, Vind I, *et al.* Environmental factors in inflammatory bowel disease: A case-control study based on a Danish inception cohort. *Journal of Crohn's and Colitis*, 2011;5(6):577–584.

[247] Haritunians T, Taylor KD, Targan SR, *et al.* Genetic predictors of medically refractory ulcerative colitis. *Inflammatory Bowel Diseases*, 2010;16(11):1830–1840.

[248] Harrell LE, Baechler EC, Behrens TW, *et al.* Mdr1, Runx3 and TGFBRii Pathways Implicated in Human IBD. *Gastroenterology*, 2006;130(4, Supplement 2):A591.

[249] Harris RA, Nagy-Szakal D, Mir SA, *et al.* DNA methylation-associated colonic mucosal immune and defense responses in treatment-naïve pediatric ulcerative colitis. *Epigenetics : official journal of the DNA Methylation Society*, 2014; 9(8):1131–7.

[250] Harris RA, Nagy-Szakal D, Pedersen N, *et al.* Genome-wide peripheral blood leukocyte DNA methylation microarrays identified a single association with inflammatory bowel diseases. *Inflammatory bowel diseases*, 2012;18(12):2334–41.

[251] Häsler R, Feng Z, Bäckdahl L, *et al.* A functional methylome map of ulcerative colitis. *Genome research*, 2012;22(11):2130–7.

[252] Hawkey CJ. Stem cells as treatment in inflammatory bowel disease. *Digestive diseases (Basel, Switzerland)*, 2012;30 Suppl 3(suppl 3):134–9.

[253] He H, Wang H, Jiao Y, *et al.* Effect of Sodium Fluoride on the Proliferation and Gene Differential Expression in Human RPMI8226 Cells. *Biological Trace Element Research*, 2015;167(1):11–17.

[254] Hedl M and Abraham C. Distinct roles for Nod2 protein and autocrine interleukin-1beta in muramyl dipeptide-induced mitogen-activated protein kinase activation and cytokine secretion in human macrophages. *The Journal of biological chemistry*, 2011;286(30):26440–9.

[255] Heikenen JB, Werlin SL, Brown CW, *et al.* Presenting symptoms and diagnostic lag in children with inflammatory bowel disease. *Inflammatory bowel diseases*, 1999;5(3):158–60.

[256] Henckaerts L, Cleynen I, Brinar M, *et al.* Genetic variation in the autophagy gene ULK1 and risk of Crohn's disease. *Inflammatory Bowel Diseases*, 2011; 17(6):1392–1397.

[257] Henderson P and Stevens C. The role of autophagy in Crohn's disease. *Cells*, 2012;1(3):492–519.

[258] Henderson P, van Limbergen JE, Wilson DC, *et al.* Genetics of childhood-onset inflammatory bowel disease. *Inflammatory bowel diseases*, 2011;17(1):346–61.

[259] Henderson P and Wilson DC. The rising incidence of paediatric-onset inflammatory bowel disease. *Archives of disease in childhood*, 2012;97(7):585–6.

[260] Henry JC, Azevedo-Pouly ACP, and Schmittgen TD. MicroRNA replacement therapy for cancer. *Pharmaceutical research*, 2011;28(12):3030–42.

[261] Heyman MB, Garnett EA, Shaikh N, *et al.* Folate concentrations in pediatric patients with newly diagnosed inflammatory bowel disease. *The American journal of clinical nutrition*, 2009;89(2):545–50.

[262] Hildebrand H, Finkel Y, Grahnquist L, *et al.* Changing pattern of paediatric inflammatory bowel disease in northern Stockholm 1990-2001. *Gut*, 2003; 52(10):1432–4.

[263] Hildebrand H, Malmborg P, Askling J, *et al.* Early-life exposures associated with antibiotic use and risk of subsequent Crohn's disease. *Scandinavian journal of gastroenterology*, 2008;43(8):961–6.

[264] Hindorff L, MacArthur J, Morales J, *et al.* A Catalog of Published Genome-Wide Association Studies.

[265] Hinske LCG, Galante PAF, Kuo WP, *et al.* A potential role for intragenic miRNAs on their hosts' interactome. *BMC genomics*, 2010;11:533.

[266] Hoefkens E, Nys K, John JM, *et al.* Genetic association and functional role of Crohn disease risk alleles involved in microbial sensing, autophagy, and endoplasmic reticulum (ER) stress. *Autophagy*, 2013;9(12):2046–2055.

[267] Hollingsworth JW, Maruoka S, Boon K, *et al.* In utero supplementation with methyl donors enhances allergic airway disease in mice. *The Journal of clinical investigation*, 2008;118(10):3462–9.

[268] Holoch D and Moazed D. RNA-mediated epigenetic regulation of gene expression. *Nature reviews. Genetics*, 2015;16(2):71–84.

[269] Hossain MZ, Healey MA, Lee C, *et al.* DNA-intercalators causing rapid re-expression of methylated and silenced genes in cancer cells. *Oncotarget*, 2013; 4(2):298–309.

[270] Houseman EA, Accomando WP, Koestler DC, *et al.* DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC bioinformatics*, 2012;13:86.

[271] Hsieh CJ, Klump B, Holzmann K, *et al.* Hypermethylation of the p16(INK4a) promoter in colectomy specimens of patients with long-standing and extensive ulcerative colitis. *Cancer Research*, 1998;58(17):3942–3945.

[272] Hsu YMS, Zhang Y, You Y, *et al.* The adaptor protein CARD9 is required for innate immune responses to intracellular pathogens. *Nature Immunology*, 2007; 8(2):198–205.

[273] Huang C, Haritunians T, Okou DT, *et al.* Characterization of Genetic Loci That Affect Susceptibility to Inflammatory Bowel Diseases in African Americans. *Gastroenterology*, 2015;149(6):1575–1586.

[274] Huang J, Ellinghaus D, Franke A, *et al.* 1000 Genomes-based imputation identifies novel and refined associations for the Wellcome Trust Case Control Consortium phase 1 Data. *European Journal of Human Genetics*, 2012;20(7):801–805.

[275] Hugot JP, Chamaillard M, Zouali H, *et al.* Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature*, 2001;411(6837):599–603.

[276] Hugot JP, Laurent-Puig P, Gower-Rousseau C, *et al.* Mapping of a susceptibility locus for Crohn's disease on chromosome 16. *Nature*, 1996;379(6568):821–3.

[277] Humphreys KJ, Conlon MA, Young GP, *et al.* Dietary manipulation of oncogenic microRNA expression in human rectal mucosa: A randomized trial. *Cancer Prevention Research*, 2014;7(8):786–795.

[278] Hur K, Niwa T, Toyoda T, *et al.* Insufficient role of cell proliferation in aberrant DNA methylation induction and involvement of specific types of inflammation. *Carcinogenesis*, 2011;32(1):35–41.

[279] Hutchins AP, Poulain S, and Miranda-Saavedra D. Genome-wide analysis of STAT3 binding in vivo predicts effectors of the anti-inflammatory response in macrophages. *Blood*, 2012;119(13):e110–9.

[280] Hutchinson JN, Raj T, Fagerness J, *et al.* Allele-specific methylation occurs at genetic variants associated with complex disease. *PloS one*, 2014;9(6):e98464.

[281] Huurre A, Kalliomäki M, Rautava S, *et al.* Mode of Delivery âĂŞ Effects on Gut Microbiota and Humoral Immunity. *Neonatology*, 2008;93(4):236–240.

[282] Hviid A, Svanström H, and Frisch M. Antibiotic use and inflammatory bowel diseases in childhood. *Gut*, 2011;60(1):49–54.

[283] Ianiro G, Bibbò S, Scaldaferri F, *et al.* Fecal Microbiota Transplantation in Inflammatory Bowel Disease. *Medicine*, 2014;93(19):e97.

[284] Imielinski M, Baldassano RN, Griffiths A, *et al.* Common variants at five new loci associated with early-onset inflammatory bowel disease. *Nature genetics*, 2009; 41(12):1335–40.

[285] Inaki K, Hillmer AM, Ukil L, *et al.* Transcriptional consequences of genomic structural aberrations in breast cancer. *Genome research*, 2011;21(5):676–87.

[286] Inohara N, Ogura Y, Fontalba A, *et al.* Host Recognition of Bacterial Muramyl Dipeptide Mediated through NOD2: Implications for Crohn's Disease. *Journal of Biological Chemistry*, 2003;278(8):5509–5512.

[287] Inoue Ki, Shiga T, and Ito Y. Runx transcription factors in neuronal development. *Neural development*, 2008;3:20.

[288] Irizarry RA, Ladd-Acosta C, Wen B, *et al.* The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nature genetics*, 2009;41(2):178–86.

[289] Issa JP, Ahuja N, Toyota M, *et al.* Accelerated age-related CpG island methylation in ulcerative colitis. *Cancer research*, 2001;61(9):3573–7.

[290] Itakura E and Mizushima N. Characterization of autophagosome formation site by a hierarchical analysis of mammalian Atg proteins. *Autophagy*, 2010;6(6):764–76.

[291] Ito H, Takazoe M, Fukuda Y, *et al.* A pilot randomized trial of a human anti-interleukin-6 receptor monoclonal antibody in active Crohn's disease. *Gastroenterology*, 2004;126(4):989–96; discussion 947.

[292] James DG, Seo DH, Chen J, *et al.* Efalizumab, a Human Monoclonal Anti-CD11a Antibody, in the Treatment of Moderate to Severe Crohn's Disease: An Open-Label Pilot Study. *Digestive Diseases and Sciences*, 2010;56(6):1806–1810.

[293] Janssen HLA, Reesink HW, Lawitz EJ, *et al.* Treatment of HCV infection by targeting microRNA. *The New England journal of medicine*, 2013;368(18):1685–94.

[294] Jess T, Frisch M, and Simonsen J. Trends in overall and cause-specific mortality among patients with inflammatory bowel disease from 1982 to 2010. *Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association*, 2013;11(1):43–8.

[295] Jiang PH, Motoo Y, Vaccaro MI, *et al.* Expression of vacuole membrane protein 1 (VMP1) in spontaneous chronic pancreatitis in the WBN/Kob rat. *Pancreas*, 2004;29(3):225–30.

[296] Jones SW, Erikson E, Blenis J, *et al.* A Xenopus ribosomal protein S6 kinase has two apparent kinase domains that are each similar to distinct protein kinases. *Proceedings of the National Academy of Sciences of the United States of America*, 1988;85(10):3377–81.

[297] Jostins L, Ripke S, Weersma RK, *et al.* Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature*, 2012;491(7422):119–24.

[298] Jounai N, Kobiyama K, Shiina M, *et al.* NLRP4 negatively regulates autophagic processes through an association with beclin1. *Journal of immunology (Baltimore, Md. : 1950)*, 2011;186(3):1646–55.

[299] Julia A, Domenech E, Ricart E, *et al.* A genome-wide association study on a southern European population identifies a new Crohn's disease susceptibility locus at RBX1-EP300. *Gut*, 2012;pp. 1440–1445.

[300] Jun JC, Cominelli F, and Abbott DW. RIP2 activity in inflammatory disease and implications for novel therapeutics. *Journal of Leukocyte Biology*, 2013; 94(5):927–932.

[301] Jurkowska RZ and Jeltsch A. Silencing of gene expression by targeted DNA methylation: concepts and approaches. *Methods in molecular biology (Clifton, N.J.)*, 2010;649:149–61.

[302] Kalla R, Ventham NT, Kennedy NA, *et al.* MicroRNAs: new players in IBD. *Gut*, 2015;64(3):504–517.

[303] Kanaan Z, Rai SN, Eichenberger MR, *et al.* Differential microRNA expression tracks neoplastic progression in inflammatory bowel disease-associated colorectal cancer. *Human mutation*, 2012;33(3):551–60.

[304] Kanaan Z, Rai SN, Eichenberger MR, *et al.* Plasma miR-21: a potential diagnostic marker of colorectal cancer. *Annals of surgery*, 2012;256(3):544–51.

[305] Kang R, Zeh HJ, Lotze MT, *et al.* The Beclin 1 network regulates autophagy and apoptosis. *Cell death and differentiation*, 2011;18(4):571–80.

[306] Kaplan GG, Hubbard J, Korzenik J, *et al.* The inflammatory bowel diseases and ambient air pollution: a novel association. *The American journal of gastroenterology*, 2010;105(11):2412–9.

[307] Kaplan GG, Jackson T, Sands BE, *et al.* The risk of developing Crohn's disease after an appendectomy: A meta-analysis. *American Journal of Gastroenterology*, 2008;103(11):2925–2931.

[308] Karatzas PS, Mantzaris GJ, Safioleas M, *et al.* DNA methylation profile of genes involved in inflammation and autoimmunity in inflammatory bowel disease. *Medicine*, 2014;93(28):e309.

[309] Kartha RV and Subramanian S. Competing endogenous RNAs (ceRNAs): new entrants to the intricacies of gene regulation. *Frontiers in genetics*, 2014;5(8).

[310] Kaser A, Lee AH, Franke A, *et al.* XBP1 links ER stress to intestinal inflammation and confers genetic risk for human inflammatory bowel disease. *Cell*, 2008; 134(5):743–756.

[311] Katoh Y and Katoh M. Hedgehog signaling, epithelial-to-mesenchymal transition and miRNA (review). *International journal of molecular medicine*, 2008; 22(3):271–5.

[312] Katsurano M, Niwa T, Yasui Y, *et al.* Early-stage formation of an epigenetic field defect in a mouse colitis model, and non-essential roles of T- and B-cells in DNA methylation induction. *Oncogene*, 2012;31(3):342–51.

[313] Kennedy NA, Adams A, Hansen R, *et al.* DOP010 Epigenome-wide analysis in childhood-onset Crohn's disease implicates MIR21 in pathogenesis and identifies multiple methylation-based diagnostic biomarkers. *Journal of Crohn's and Colitis*, 2014;8:S19.

[314] Kenny EE, Pe'er I, Karban A, *et al.* A genome-wide scan of ashkenazi jewish crohn's disease suggests novel susceptibility loci. *PLoS Genetics*, 2012;8(3).

[315] Khalili H, Ananthakrishnan AN, Higuchi LM, *et al.* Measures of Adiposity and Risk of Crohn's Disease and Ulcerative Colitis. *Gastroenterology*, 2013;144(5 (Supp)):S–48.

[316] Khalili H, Ananthakrishnan AN, Konijeti GG, *et al.* Physical activity and risk of inflammatory bowel disease : prospective study from the Nurses ' Health Study cohorts. *British Medical Journal*, 2013;347(November):f6633.

[317] Khalili H, Higuchi LM, Ananthakrishnan AN, *et al.* Oral contraceptives, reproductive factors and risk of inflammatory bowel disease. *Gut*, 2013;62(8):1153–9.

[318] Khan KJ, Ullman TA, Ford AC, *et al.* Antibiotic therapy in inflammatory bowel disease: a systematic review and meta-analysis. *The American journal of gastroenterology*, 2011;106(4):661–673.

[319] Khor B, Gardet A, and Xavier RJ. Genetics and pathogenesis of inflammatory bowel disease. *Nature*, 2011;474(7351):307–17.

[320] Kile BT, Schulman BA, Alexander WS, *et al.* The SOCS box: a tale of destruction and degradation. *Trends in biochemical sciences*, 2002;27(5):235–41.

[321] Kim B, Sasaki Y, and Egawa T. Restriction of Nonpermissive RUNX3 Protein Expression in T Lymphocytes by the Kozak Sequence. *Journal of immunology (Baltimore, Md. : 1950)*, 2015;195(4):1517–23.

[322] Kim KA, Lin W, Tai AW, *et al.* Hepatic SOCS3 expression is strongly associated with non-response to therapy and race in HCV and HCV/HIV infection. *Journal of hepatology*, 2009;50(4):705–11.

[323] Kim KW, Kim SH, Lee EY, *et al.* Extracellular Signal-regulated Kinase/90-kDa Ribosomal S6 Kinase/Nuclear Factor-kB Pathway Mediates Phorbol 12-Myristate 13-Acetate-induced Megakaryocytic Differentiation of K562 Cells. *Journal of Biological Chemistry*, 2001;276(16):13186–13191.

[324] Kim SC and Ferry GD. Inflammatory bowel diseases in pediatric and adolescent patients: Clinical, therapeutic, and psychosocial considerations. *Gastroenterology*, 2004;126(6):1550–1560.

[325] Kim TD, Lee SUH, Yun S, *et al.* Human microRNA-27a * targets Prf1 and GzmB expression to regulate NK cell cytotoxicity. *Blood*, 2011;118(20):5476–5486.

[326] Kim TO, Park J, Kang MJ, *et al.* DNA hypermethylation of a selective gene panel as a risk marker for colon cancer in patients with ulcerative colitis. *International journal of molecular medicine*, 2013;31(5):1255–61.

[327] Kim VN, Han J, and Siomi MC. Biogenesis of small RNAs in animals. *Nature reviews. Molecular cell biology*, 2009;10(2):126–39.

[328] Kim YI. Folate: a magic bullet or a double edged sword for colorectal cancer prevention? *Gut*, 2006;55(10):1387–9.

[329] Kim YJ, Hwang SJ, Bae YC, *et al.* MiR-21 regulates adipogenic differentiation through the modulation of TGF-beta signaling in mesenchymal stem cells derived from human adipose tissue. *Stem cells (Dayton, Ohio)*, 2009;27(12):3093–102.

[330] Kirsner JB. Historical origins of current IBD concepts. *World journal of gastroenterology : WJG*, 2001;7(2):175–84.

[331] Kishi-Itakura C, Koyama-Honda I, Itakura E, *et al.* Ultrastructural analysis of autophagosome organization using mammalian autophagy-deficient cells. *Journal of Cell Science*, 2014;127(22):4984–4984.

[332] Kisiel JB, Yab TC, Nazer Hussain FT, *et al.* Stool DNA testing for the detection of colorectal neoplasia in patients with inflammatory bowel disease. *Alimentary pharmacology & therapeutics*, 2013;37(5):546–54.

[333] Kitis G, Thompson H, and Allan RN. Finger clubbing in inflammatory bowel disease: its prevalence and pathogenesis. *British medical journal*, 1979;2(6194):825–8.

[334] Klement E, Lysy J, Hoshen M, *et al.* Childhood hygiene is associated with the risk for inflammatory bowel disease: a population-based study. *The American journal of gastroenterology*, 2008;103(7):1775–82.

[335] Klement E and Reif S. Breastfeeding and risk of inflammatory bowel disease. *The American journal of clinical nutrition*, 2005;82:486.

[336] Koenig JE, Spor A, Scalfone N, *et al.* Succession of microbial consortia in the developing infant gut microbiome. *Proceedings of the National Academy of Sciences of the United States of America*, 2011;108 Suppl:4578–4585.

[337] Koizumi K, Alonso S, Miyaki Y, *et al.* Array-based identification of common DNA methylation alterations in ulcerative colitis. *International journal of oncology*, 2012;40(4):983–94.

[338] Kosumi K, Baba Y, Ishimoto T, *et al.* Relationship between LINE-1 hypomethylation and Helicobacter pylori infection in gastric mucosae. *Medical Oncology*, 2015;32(4).

[339] Kotlarz D, Beier R, Murugan D, *et al.* Loss of Interleukin-10 Signaling and Infantile Inflammatory Bowel Disease: Implications for Diagnosis and Therapy. *Gastroenterology*, 2012;143(2):347–355.

[340] Koukos G, Polytarchou C, Kaplan JL, *et al.* MicroRNA-124 regulates STAT3 expression and is down-regulated in colon tissues of pediatric patients with ulcerative colitis. *Gastroenterology*, 2013;145(4):842–52.

[341] Koutroubakis IE and Vlachonikolis IG. Appendectomy and the development of ulcerative colitis: Results of a metaanalysis of published case-control studies. *American Journal of Gastroenterology*, 2000;95(1):171–176.

[342] Koutroumpakis E, Ramos-Rivers C, Regueiro M, *et al.* Association Between Long-Term Lipid Profiles and Disease Severity in a Large Cohort of Patients with Inflammatory Bowel Disease. *Digestive Diseases and Sciences*, 2015;.

[343] Kovarik JJ, Tillinger W, Hofer J, *et al.* Impaired anti-inflammatory efficacy of n-butyrate in patients with IBD. *European journal of clinical investigation*, 2011; 41(3):291–8.

[344] Kozomara A and Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic acids research*, 2014;42(Database issue):D68–73.

[345] Kraiczy J, Nayak K, Ross A, *et al.* Assessing DNA methylation in the developing human intestinal epithelium: potential link to inflammatory bowel disease. *Mucosal Immunology*, 2015;(May):1–12.

[346] Kriaucionis S and Heintz N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science*, 2009;324(5929):929–30.

[347] Krichevsky AM and Gabriely G. miR-21: a small multi-faceted RNA. *Journal of cellular and molecular medicine*, 2009;13(1):39–53.

[348] Kuenzel S, Till A, Winkler M, *et al.* The Nucleotide-Binding Oligomerization Domain-Like Receptor NLRC5 Is Involved in IFN-Dependent Antiviral Immune Responses. *The Journal of Immunology*, 2010;184(4):1990–2000.

[349] Kugathasan S, Baldassano RN, Bradfield JP, *et al.* Loci on 20q13 and 21q22 are associated with pediatric-onset inflammatory bowel disease. *Nature genetics*, 2008;40(10):1211–1215.

[350] Kundaje A, Meuleman W, Ernst J, *et al.* Integrative analysis of 111 reference human epigenomes. *Nature*, 2015;518(7539):317–330.

[351] Kyme P, Thoennissen NH, Tseng CW, *et al.* C/EBPϵ mediates nicotinamide-enhanced clearance of Staphylococcus aureus in mice. *The Journal of clinical investigation*, 2012;122(9):3316–29.

[352] Ladewig E, Okamura K, Flynt AS, *et al.* Discovery of hundreds of mirtrons in mouse and human small RNA data. *Genome research*, 2012;22(9):1634–45.

[353] Laird PW. Principles and challenges of genomewide DNA methylation analysis. *Nature Reviews Genetics*, 2010;11(3):191–203.

[354] Lakatos PL, Szamosi T, and Lakatos L. Smoking in inflammatory bowel diseases: good, bad or ugly? *World journal of gastroenterology : WJG*, 2007;13(46):6134–9.

[355] Lamkanfi M and Kanneganti TD. Regulation of immune pathways by the NOD-like receptor NLRC5. *Immunobiology*, 2012;217(1):13–16.

[356] Lang R, Pauleau AL, Parganas E, *et al.* SOCS3 regulates the plasticity of gp130 signaling. *Nature immunology*, 2003;4(6):546–550.

[357] Lango Allen H, Estrada K, Lettre G, *et al.* Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature*, 2010; 467(7317):832–8.

[358] Lapidus A. Crohn's disease in Stockholm County during 1990-2001: an epidemiological update. *World journal of gastroenterology : WJG*, 2006;12(1):75–81.

[359] Lara R, Seckl MJ, and Pardo OE. The p90 RSK family members: Common functions and isoform specificity. *Cancer Research*, 2013;73(17):5301–5308.

[360] Larsen S, Bendtzen K, and Nielsen OH. Extraintestinal manifestations of inflammatory bowel disease: epidemiology, diagnosis, and management. *Annals of medicine*, 2010;42(2):97–114.

[361] Lawrance IC. Novel topical therapies for distal colitis. *World journal of gastrointestinal pharmacology and therapeutics*, 2010;1(5):87–93.

[362] Lawson M, Uciechowska U, Schemies J, *et al.* Inhibitors to understand molecular mechanisms of NAD(+)-dependent deacetylases (sirtuins). *Biochimica et biophysica acta*, 2010;1799(10-12):726–39.

[363] Lee KH, Biswas A, Liu YJ, *et al.* Proteasomal degradation of Nod2 mediates tolerance to bacterial cell wall components. *Journal of Biological Chemistry*, 2012;287(47):39800–39811.

[364] Lee SA, Tsao T, Yang KC, *et al.* Construction and analysis of the protein-protein interaction networks for schizophrenia, bipolar disorder, and major depression. *BMC Bioinformatics*, 2011;12(Suppl 13):S20.

[365] Leek JT and Storey JD. Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS genetics*, 2007;3(9):1724–35.

[366] Lees CW, Zacharias WJ, Tremelling M, *et al.* Analysis of germline GLI1 variation implicates hedgehog signalling in the regulation of intestinal inflammatory pathways. *PLoS Medicine*, 2008;5(12):1761–1775.

[367] Leighton IA, Dalby KN, Barry Caudwell F, *et al.* Comparison of the specificities of p70 S6 kinase and MAPKAP kinase-1 identifies a relatively specific substrate for p70 S6 kinase: the N-terminal kinase domain of MAPKAP kinase-1 is essential for peptide phosphorylation. *FEBS Letters*, 1995;375(3):289–293.

[368] Leoni F, Zaliani A, Bertolini G, *et al.* The antitumor histone deacetylase inhibitor suberoylanilide hydroxamic acid exhibits antiinflammatory properties via suppression of cytokines. *Proceedings of the National Academy of Sciences of the United States of America*, 2002;99(5):2995–3000.

[369] Leung CH, Lam W, Ma DL, *et al.* Butyrate mediates nucleotide-binding and oligomerisation domain (NOD) 2-dependent mucosal immune responses against peptidoglycan. *European journal of immunology*, 2009;39(12):3529–37.

[370] Levine A, Griffiths A, Markowitz J, *et al.* Pediatric modification of the Montreal classification for inflammatory bowel disease. *Inflammatory Bowel Diseases*, 2011; 17(6):1314–1321.

[371] Levine A, Kugathasan S, Annese V, *et al.* Pediatric onset Crohn's colitis is characterized by genotype-dependent age-related susceptibility. *1509-1515*, 2007; 13(12):1509–15.

[372] Levine B, Mizushima N, and Virgin HW. Autophagy in immunity and inflammation. *Nature*, 2011;469(7330):323–335.

[373] Levine J and Burakoff R. Extraintestinal manifestations of inflammatory bowel disease. *Gastroenterology and Hepatology*, 2011;7(4):235–241.

[374] Lewis BP, Burge CB, and Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*, 2005;120(1):15–20.

[375] Li W, Li J, and Bao J. Microautophagy: lesser-known self-eating. *Cellular and Molecular Life Sciences*, 2012;69(7):1125–1136.

[376] Li Y, de Haar C, Chen M, *et al.* Disease-related expression of the IL6/STAT3/SOCS3 signalling pathway in ulcerative colitis and ulcerative colitis-related carcinogenesis. *Gut*, 2010;59(2):227–235.

[377] Libioulle C, Louis E, Hansoul S, *et al.* Novel Crohn disease locus identified by genome-wide association maps to a gene desert on 5p13.1 and modulates expression of PTGER4. *PLoS Genetics*, 2007;3(4):0538–0543.

[378] Libri V, Helwak A, Miesen P, *et al.* Murine cytomegalovirus encodes a miR-27 inhibitor disguised as a target. *Proceedings of the National Academy of Sciences of the United States of America*, 2012;109(1):279–84.

[379] Lin J, Haffner MC, Zhang Y, *et al.* Disulfiram is a DNA demethylating agent and inhibits prostate cancer cell growth. *The Prostate*, 2011;71(4):333–43.

[380] Lin Z, Hegarty JP, Cappel JA, *et al.* Identification of disease-associated DNA methylation in intestinal tissues from patients with inflammatory bowel disease. *Clinical genetics*, 2011;80(1):59–67.

[381] Lin Z, Hegarty JP, Yu W, *et al.* Identification of disease-associated DNA methylation in B cells from Crohn's disease and ulcerative colitis patients. *Digestive diseases and sciences*, 2012;57(12):3145–53.

[382] Liu D, Sheng C, Gao S, *et al.* SOCS3 Drives Proteasomal Degradation of TBK1 and Negatively Regulates Antiviral Innate Immunity. *Molecular and cellular biology*, 2015;35(14):2400–2413.

[383] Liu J, Xie YS, Wang FL, *et al.* Cytotoxicity of 5-Aza-2'-deoxycytidine against gastric cancer involves DNA damage in an ATM-P53 dependent signaling pathway and demethylation of P16INK4A. *Biomedicine and Pharmacotherapy*, 2013; 67(1):78–87.

[384] Liu JZ, van Sommeren S, Huang H, *et al.* Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nature Genetics*, 2015;47(9):979–989.

[385] Liu Q, Shimoyama T, Suzuki K, *et al.* Effect of sodium butyrate on reactive oxygen species generation by human neutrophils. *Scandinavian journal of gastroenterology*, 2001;36(7):744–750.

[386] Liu Z, Chen L, Zhang X, *et al.* RUNX3 regulates vimentin expression via miR-30a during epithelial-mesenchymal transition in gastric cancer cells. *Journal of Cellular and Molecular Medicine*, 2014;18(4):610–623.

[387] Liu Z, Lee J, Krummey S, *et al.* The kinase LRRK2 is a regulator of the transcription factor NFAT that modulates the severity of inflammatory. *Nature immunology*, 2011;12(11):1063–70.

[388] Liverani E. Mycobacterium avium subspecies paratuberculosis in the etiology of Crohn's disease, cause or epiphenomenon? *World Journal of Gastroenterology*, 2014;20(36):13060.

[389] Lo Sasso G, Ryu D, Mouchiroud L, *et al.* Loss of Sirt1 function improves intestinal anti-bacterial defense and protects from colitis-induced colorectal cancer. *PloS one*, 2014;9(7):e102495.

[390] Lobatón T, Azuara D, Rodríguez-Moranta F, *et al.* Relationship between methylation and colonic inflammation in inflammatory bowel disease. *World journal of gastroenterology*, 2014;20(30):10591–8.

[391] Löffler D, Brocke-Heidrich K, Pfeifer G, *et al.* Interleukin-6 dependent survival of multiple myeloma cells involves the Stat3-mediated induction of microRNA-21 through a highly conserved enhancer. *Blood*, 2007;110(4):1330–3.

[392] López-Pedrera C, Pérez-Sánchez C, Ramos-Casals M, *et al.* Cardiovascular Risk in Systemic Autoimmune Diseases: Epigenetic Mechanisms of Immune Regulatory Functions. *Clinical & developmental immunology*, 2012;2012:974648.

[393] Lu C, Chen J, Xu HG, *et al.* MIR106B and MIR93 Prevent Removal of Bacteria From Epithelial Cells by Disrupting ATG16L1-Mediated Autophagy. *Gastroenterology*, 2014;146(1):188–99.

[394] Lu TX, Hartner J, Lim EJ, *et al.* MicroRNA-21 limits in vivo immune response-mediated activation of the IL-12/IFN-gamma pathway, Th1 polarization, and the severity of delayed-type hypersensitivity. *Journal of immunology (Baltimore, Md. : 1950)*, 2011;187(6):3362–73.

[395] Lu WG, Zou YF, Feng XL, *et al.* Association of NOD1 (CARD4) insertion/deletion polymorphism with susceptibility to IBD: A meta-analysis. *World journal of gastroenterology : WJG*, 2010;16(34):4348.

[396] Ludwig K, Fassan M, Mescoli C, *et al.* PDCD4/miR-21 dysregulation in inflammatory bowel disease-associated carcinogenesis. *Virchows Archiv : an international journal of pathology*, 2013;462(1):57–63.

[397] Luger K, Mäder AW, Richmond RK, *et al.* Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature*, 1997;389(6648):251–60.

[398] Lührs H, Gerke T, Müller JG, *et al.* Butyrate inhibits NF-kappaB activation in lamina propria macrophages of patients with ulcerative colitis. *Scandinavian journal of gastroenterology*, 2002;37(4):458–66.

[399] Lum JJ, DeBerardinis RJ, and Thompson CB. Autophagy in metazoans: cell survival in the land of plenty. *Nature Reviews Molecular Cell Biology*, 2005; 6(6):439–448.

[400] Ma X, Ezzeldin HH, and Diasio RB. Histone deacetylase inhibitors: current status and overview of recent clinical trials. *Drugs*, 2009;69(14):1911–34.

[401] Mackner LM, Greenley RN, Szigethy E, *et al.* Clinical Report of the North American Society for Pediatric. *Journal of Pediatric Gastroenterology and Nutrition*, 2013;56(4):449–458.

[402] Mahid SS, Minor KS, Soto RE, *et al.* Smoking and inflammatory bowel disease: a meta-analysis. *Mayo Clinic proceedings*, 2006;81(11):1462–71.

[403] Major EO. Progressive multifocal leukoencephalopathy in patients on immunomodulatory therapies. *Annual review of medicine*, 2010;61:35–47.

[404] Majumdar A, Nagaraj R, and Banerjee U. strawberry notch encodes a conserved nuclear protein that functions downstream of Notch and regulates gene expression along the developing wing margin of Drosophila. *Genes & Development*, 1997; 11(10):1341–1353.

[405] Manolio TA, Collins FS, Cox NJ, *et al.* Finding the missing heritability of complex diseases. *Nature*, 2009;461(7265):747–753.

[406] Marabita F, Almgren M, Lindholm ME, *et al.* An evaluation of analysis pipelines for DNA methylation profiling using the Illumina HumanMethylation450 BeadChip platform. *Epigenetics : official journal of the DNA Methylation Society*, 2013;8(3):333–46.

[407] Marcinowski L, Tanguy M, Krmpotic A, *et al.* Degradation of cellular mir-27 by a novel, highly abundant viral transcript is important for efficient virus replication in vivo. *PLoS pathogens*, 2012;8(2):e1002510.

[408] Markowitz J, Kugathasan S, Dubinsky M, *et al.* Age of diagnosis influences serologic responses in children with Crohn's disease: a possible clue to etiology? *Inflammatory bowel diseases*, 2009;15(5):714–9.

[409] Marks PA, Richon VM, and Rifkind RA. Histone Deacetylase Inhibitors: Inducers of Differentiation or Apoptosis of Transformed Cells. *JNCI Journal of the National Cancer Institute*, 2000;92(15):1210–1216.

[410] Marlow GJ, van Gent D, and Ferguson LR. Why interleukin-10 supplementation does not work in Crohn's disease patients. *World journal of gastroenterology : WJG*, 2013;19(25):3931–41.

[411] Marson A, Levine SS, Cole MF, *et al.* Connecting microRNA Genes to the Core Transcriptional Regulatory Circuitry of Embryonic Stem Cells. *Cell*, 2008; 134(3):521–533.

[412] Martinez-Medina M and Garcia-Gil LJ. Escherichia coli in chronic inflammatory bowel diseases: An update on adherent invasive Escherichia coli pathogenicity. *World journal of gastrointestinal pathophysiology*, 2014;5(3):213–27.

[413] Maruyama K, Uematsu S, Kondo T, *et al.* Strawberry notch homologue 2 regulates osteoclast fusion by enhancing the expression of DC-STAMP. *The Journal of experimental medicine*, 2013;210(10):1947–60.

[414] Marzano AV, Ishak RS, Saibeni S, *et al.* Autoinflammatory skin disorders in inflammatory bowel diseases, pyoderma gangrenosum and Sweet's syndrome: a comprehensive review and disease classification criteria. *Clinical reviews in allergy & immunology*, 2013;45:202–10.

[415] Mathelier A, Fornes O, Arenillas DJ, *et al.* JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic acids research*, 2015;.

[416] Mathern DR, Laitman LE, Hovhannisyan Z, *et al.* Mouse and human Notch-1 regulate mucosal immune responses. *Mucosal immunology*, 2014;7(4):995–1005.

[417] Matsuoka K and Kanai T. The gut microbiota and inflammatory bowel disease. *Seminars in Immunopathology*, 2015;37(1):47–55.

[418] Maunakea AK, Nagarajan RP, Bilenky M, *et al.* Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature*, 2010;466(7303):253–7.

[419] Mayberry MK, Probert C, Srivastava E, *et al.* Perceived discrimination in education and employment by people with Crohn's disease: a case control study of educational achievement and employment. *Gut*, 1992;33(3):312–4.

[420] McDermott E, Ryan EJ, Tosetto M, *et al.* DNA methylation profiling in inflammatory bowel disease provides new insights into disease pathogenesis. *Journal of Crohn's & colitis*, 2015;pp. 1–33.

[421] McGovern DPB, Butler H, Ahmad T, *et al.* TUCAN (CARD8) Genetic Variants and Inflammatory Bowel Disease. *Gastroenterology*, 2006;131(4):1190–1196.

[422] McGovern DPB, Gardet A, Törkvist L, *et al.* Genome-wide association identifies multiple ulcerative colitis susceptibility loci. *Nature Genetics*, 2010;42(4):332–337.

[423] McGovern DPB, Jones MR, Taylor KD, *et al.* Fucosyltransferase 2 (FUT2) non-secretor status is associated with Crohn's disease. *Human Molecular Genetics*, 2010;19(17):3468–3476.

[424] McIlroy D, Meyer L, Dudoit Y, *et al.* Polymorphism in the proximal promoter region of the perforin gene and its impact on the course of HIV infection. *International Journal of Immunogenetics*, 2006;33(2):73–79.

[425] Meinzer U, Ideström M, Alberti C, *et al.* Ileal involvement is age dependent in pediatric Crohn's disease. *Inflammatory bowel diseases*, 2005;11(7):639–644.

[426] Meissner TB, Li A, Biswas A, *et al.* NLR family member NLRC5 is a transcriptional regulator of MHC class I genes. *Proceedings of the National Academy of Sciences of the United States of America*, 2010;107(31):13794–13799.

[427] Meissner TB, Liu YJ, Lee KH, *et al.* NLRC5 cooperates with the RFX transcription factor complex to induce MHC class I gene expression. *Journal of immunology (Baltimore, Md. : 1950)*, 2012;188(10):4951–8.

[428] Meister G. Argonaute proteins: functional insights and emerging roles. *Nature reviews. Genetics*, 2013;14(7):447–59.

[429] Mikhailova TL, Sishkova E, Poniewierka E, *et al.* Randomised clinical trial: the efficacy and safety of propionyl-L-carnitine therapy in patients with ulcerative colitis receiving stable oral treatment. *Alimentary pharmacology & therapeutics*, 2011;34(9):1088–97.

[430] Mikkelsen TS, Xu Z, Zhang X, *et al.* Comparative epigenomic analysis of murine and human adipogenesis. *Cell*, 2010;143(1):156–69.

[431] Minocha A and Raczkowski CA. Role of appendectomy and tonsillectomy in pathogenesis of ulcerative colitis. *Digestive Diseases and Sciences*, 1997; 42(7):1567–1569.

[432] Mo J, Boyle JP, Howard CB, *et al.* Pathogen sensing by nucleotide-binding oligomerization domain-containing protein 2 (NOD2) is mediated by direct binding to muramyl dipeptide and ATP. *Journal of Biological Chemistry*, 2012; 287(27):23057–23067.

[433] Mojarad EN, Kuppen PJK, Aghdaei HA, *et al.* The CpG island methylator phenotype (CIMP) in colorectal cancer. *Gastroenterol Hepatol Bed Bench*, 2013; 6(3):120–128.

[434] Mokry M, Middendorp S, Wiegerinck CL, *et al.* Many inflammatory bowel disease risk loci include regions that regulate gene expression in immune cells and the intestinal epithelium. *Gastroenterology*, 2014;146(4):1040–7.

[435] Molejon MI, Ropolo A, Lo Re A, *et al.* The VMP1-Beclin 1 interaction regulates autophagy induction. *Scientific reports*, 2013;3:1055.

[436] Molodecky NA, Soon IS, Rabi DM, *et al.* Increasing incidence and prevalence of the inflammatory bowel diseases with time, based on systematic review. *Gastroenterology*, 2012;142(1):46–54.

[437] Monteleone G, Neurath MF, Ardizzone S, *et al.* Mongersen, an oral SMAD7 antisense oligonucleotide, and Crohn's disease. *The New England journal of medicine*, 2015;372(12):1104–13.

[438] Moran GW, Dubeau MF, Kaplan GG, *et al.* The increasing weight of Crohn's disease subjects in clinical trials: a hypothesis-generatings time-trend analysis. *Inflammatory bowel diseases*, 2013;19(13):2949–56.

[439] Morgan AR, Lam WJ, Han DY, *et al.* Association Analysis of ULK1 with Crohn's Disease in a New Zealand Population. *Gastroenterology research and practice*, 2012;2012:715309.

[440] Morís G. Inflammatory bowel disease: an increased risk factor for neurologic complications. *World journal of gastroenterology : WJG*, 2014;20(5):1228–37.

[441] Moriyama T, Matsumoto T, Nakamura S, *et al.* Hypermethylation of p14 (ARF) may be predictive of colitic cancer in patients with ulcerative colitis. *Diseases of the colon and rectum*, 2007;50(9):1384–92.

[442] Morris DL, Montgomery SM, Thompson NP, *et al.* Measles vaccination and inflammatory bowel disease: a national British Cohort Study. *The American journal of gastroenterology*, 2000;95(12):3507–12.

[443] Morris TJ, Butcher LM, Feber A, *et al.* ChAMP: 450k Chip Analysis Methylation Pipeline. *Bioinformatics (Oxford, England)*, 2014;30(3):428–30.

[444] Mótyán J, Bagossi P, BenkÅŚ S, *et al.* A molecular model of the full-length human NOD-like receptor family CARD domain containing 5 (NLRC5) protein. *BMC Bioinformatics*, 2013;14(1):275.

[445] Mowat C, Cole A, Windsor A, *et al.* Guidelines for the management of inflammatory bowel disease in adults. *Gut*, 2011;60(5):571–607.

[446] Munyaka PM, Khafipour E, and Ghia JE. External Influence of Early Childhood Establishment of Gut Microbiota and Subsequent Health Implications. *Frontiers in Pediatrics*, 2014;2(October):1–9.

[447] Münzel M, Globisch D, and Carell T. 5-Hydroxymethylcytosine, the sixth base of the genome. *Angewandte Chemie (International ed. in English)*, 2011; 50(29):6460–8.

[448] Murdaca G, Colombo BM, Cagnati P, *et al.* Endothelial dysfunction in rheumatic autoimmune diseases. *Atherosclerosis*, 2012;224(2):309–317.

[449] Murphy LO, Smith S, Chen RH, *et al.* Molecular interpretation of ERK signal duration by immediate early gene products. *Nature cell biology*, 2002;4(8):556–564.

[450] Murray PJ. Understanding and exploiting the endogenous interleukin-10/STAT3-mediated anti-inflammatory response. *Current Opinion in Pharmacology*, 2006; 6(Il):379–386.

[451] Na SY, Park SS, and Seo JK. Genetic Polymorphisms in Autophagy-Associated Genes in Korean Children With Early-Onset Crohn Disease. *Journal of pediatric gastroenterology and nutrition*, 2015;61(3):285–91.

[452] Nakamura N, Lill JR, Phung Q, *et al.* Endosomes are specialized platforms for bacterial sensing and NOD2 signalling. *Nature*, 2014;509(7499):240–4.

[453] National Institute for Health and Care Excellence. Crohn' s disease: management, 2012.

[454] Neerincx A, Lautz K, Menning M, *et al.* A role for the human nucleotide-binding domain, leucine-rich repeat-containing family member NLRC5 in antiviral responses. *J Biol Chem*, 2010;285(34):26223–26232.

[455] Ng SC, Bernstein CN, Vatn MH, *et al.* Geographical variability and environmental risk factors in inflammatory bowel disease. *Gut*, 2013;62(4):630–49.

[456] Nguyen HTT, Dalmasso G, Müller S, *et al.* Crohn's Disease-Associated Adherent Invasive Escherichia coli Modulate Levels of microRNAs in Intestinal Epithelial Cells to Reduce Autophagy. *Gastroenterology*, 2014;146(2):508–19.

[457] Nielsen BS, Jørgensen S, Fog JU, *et al.* High levels of microRNA-21 in the stroma of colorectal cancers predict short disease-free survival in stage II colon cancer patients. *Clinical & experimental metastasis*, 2011;28(1):27–38.

[458] Nimmo ER, Prendergast JG, Aldhous MC, *et al.* Genome-wide methylation profiling in Crohn's disease identifies altered epigenetic regulation of key host defense mechanisms including the Th17 pathway. *Inflammatory bowel diseases*, 2012;18(5):889–99.

[459] Nimmo ER, Stevens C, Phillips AM, *et al.* TLE1 modifies the effects of NOD2 in the pathogenesis of Crohn's disease. *Gastroenterology*, 2011;141(3):972–981.e1–2.

[460] Niwa T, Tsukamoto T, Toyoda T, *et al.* Inflammatory processes triggered by Helicobacter pylori infection cause aberrant DNA methylation in gastric epithelial cells. *Cancer Research*, 2010;70(4):1430–1440.

[461] Noble CL, Abbas AR, Cornelius J, *et al.* Regional variation in gene expression in the healthy colon is dysregulated in ulcerative colitis. *Gut*, 2008;57(10):1398–405.

[462] Noble CL, Abbas AR, Lees CW, *et al.* Characterization of intestinal gene expression profiles in Crohn's disease by genome-wide microarray analysis. *Inflammatory bowel diseases*, 2010;16(10):1717–28.

[463] Noguchi E, Homma Y, Kang X, *et al.* A Crohn's diseaseâĂŞassociated NOD2 mutation suppresses transcription of human IL10 by inhibiting activity of the nuclear ribonucleoprotein hnRNP-A1. *Nature Immunology*, 2009;10(5):471–479.

[464] Nowak J, Archange C, Tardivel-Lacombe J, *et al.* The TP53INP2 protein is required for autophagy in mammalian cells. *Molecular biology of the cell*, 2009; 20(3):870–81.

[465] Ntziachristos P, Lim JS, Sage J, *et al.* From fly wings to targeted cancer therapies: A centennial for notch signaling. *Cancer Cell*, 2014;25(3):318–334.

[466] Ogura Y, Bonen DK, Inohara N, *et al.* A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature*, 2001;411(6837):603–6.

[467] Ogura Y, Inohara N, Benito A, *et al.* Nod2, a Nod1/Apaf-1 Family Member That Is Restricted to Monocytes and Activates NF-κB. *Journal of Biological Chemistry*, 2001;276(7):4812–4818.

[468] Okada Y, Yamazaki K, Umeno J, *et al.* HLA-Cw*1202-B *5201-DRB1 *1502 haplotype increases risk for ulcerative colitis but reduces risk for crohn's disease. *Gastroenterology*, 2011;141(3):864–871.e5.

[469] Okou DT and Kugathasan S. Role of Genetics in Pediatric Inflammatory Bowel Disease. *Inflammatory Bowel Diseases*, 2014;20(10):1878–1884.

[470] Olejniczak M, Galka P, and Krzyzosiak WJ. Sequence-non-specific effects of RNA interference triggers and microRNA regulators. *Nucleic Acids Research*, 2010;38(1):1–16.

[471] Omrane I and Benammar-Elgaaied A. The immune microenvironment of the colorectal tumor: Involvement of immunity genes and microRNAs belonging to the TH17 pathway. *Biochimica et Biophysica Acta - Reviews on Cancer*, 2015; 1856(1):28–38.

[472] Orenstein SJ and Cuervo AM. Chaperone-mediated autophagy: molecular mechanisms and physiological relevance. *Seminars in cell & developmental biology*, 2010;21(7):719–26.

[473] Orta ML, Calderón-Montaño JM, Domínguez I, *et al.* 5-Aza-2'-deoxycytidine causes replication lesions that require Fanconi anemia-dependent homologous recombination for repair. *Nucleic Acids Research*, 2013;41(11):5827–5836.

[474] Ott C and Schölmerich J. Extraintestinal manifestations and complications in IBD. *Nature Reviews Gastroenterology & Hepatology*, 2013;10(10):585–595.

[475] Otto F, Lübbert M, and Stock M. Upstream and downstream targets of RUNX proteins. *Journal of Cellular Biochemistry*, 2003;89(1):9–18.

[476] Ozsolak F, Poling LL, Wang Z, *et al.* Chromatin structure analyses identify miRNA promoters. *Genes & development*, 2008;22(22):3172–83.

[477] Pages H. BSgenome: Infrastructure for Biostrings-based genome data packages and support for efficient SNP representation.

[478] Pan W, Zhu S, Yuan M, *et al.* MicroRNA-21 and microRNA-148a contribute to DNA hypomethylation in lupus CD4+ T cells by directly and indirectly targeting DNA methyltransferase 1. *Journal of immunology*, 2010;184(12):6773–6781.

[479] Panta GR, Kaur S, Cavin LG, *et al.* ATM and the catalytic subunit of DNA-dependent protein kinase activate NF-kappaB through a common MEK/extracellular signal-regulated kinase/p90(rsk) signaling pathway in response to distinct forms of DNA damage. *Molecular and cellular biology*, 2004; 24(5):1823–35.

[480] Papa E, Docktor M, Smillie C, *et al.* Non-Invasive Mapping of the Gastrointestinal Microbiota Identifies Children with Inflammatory Bowel Disease. *PLoS ONE*, 2012;7(6):e39242.

[481] Paraskevi A, Theodoropoulos G, Papaconstantinou I, *et al.* Circulating MicroRNA in inflammatory bowel disease. *Journal of Crohn's and Colitis*, 2012; 6(9):900–904.

[482] Park ES, Kim H, Suh JM, *et al.* Thyrotropin induces SOCS-1 (suppressor of cytokine signaling-1) and SOCS-3 in FRTL-5 thyroid cells. *Molecular endocrinology (Baltimore, Md.)*, 2000;14(3):440–8.

[483] Park Y and Wu H. Differential methylation analysis for BS-seq data under general experimental design. *Bioinformatics*, 2016;32(10):1446–1453.

[484] Parker W. The "hygiene hypothesis" for allergic disease is a misnomer. *Bmj*, 2014;349:g5267.

[485] Parkes GC, Whelan K, and Lindsay JO. Smoking in inflammatory bowel disease: impact on disease course and insights into the aetiology of its effect. *Journal of Crohn's & colitis*, 2014;8(8):717–25.

[486] Parkes M, Barrett JC, Prescott NJ, *et al.* Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility. *Nature genetics*, 2007;39(7):830–832.

[487] Patel S, Behara R, Swanson G, *et al.* Alcohol and the Intestine. *Biomolecules*, 2015;5(4):2573–2588.

[488] Pearce LR, Komander D, and Alessi DR. The nuts and bolts of AGC protein kinases. *Nature reviews. Molecular cell biology*, 2010;11(1):9–22.

[489] Penders J, Gerhold K, Thijs C, *et al.* New insights into the hygiene hypothesis in allergic diseases: mediation of sibling and birth mode effects by the gut microbiota. *Gut microbes*, 2014;5(2):239–44.

[490] Philpott DJ and Girardin SE. Nod-like receptors: Sentinels at host membranes. *Current Opinion in Immunology*, 2010;22(4):428–434.

[491] Philpott DJ, Sorbara MT, Robertson SJ, *et al.* NOD proteins: regulators of inflammation in health and disease. *Nature reviews. Immunology*, 2014;14(1):9–23.

[492] Picascia A, Grimaldi V, Pignalosa O, *et al.* Epigenetic control of Autoimmune Diseases: From Bench to Bedside. *Clinical Immunology*, 2015;157(1):1–15.

[493] Pidsley R, Y Wong CC, Volta M, *et al.* A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC genomics*, 2013;14:293.

[494] Pigneur B, Seksik P, Viola S, *et al.* Natural history of Crohn's disease: comparison between childhood- and adult-onset disease. *Inflammatory bowel diseases*, 2010; 16(6):953–61.

[495] Pineton de Chambrun G, Dauchet L, Gower-Rousseau C, *et al.* Vaccination and Risk for Developing Inflammatory Bowel Disease: A Meta-Analysis of Case-Control and Cohort Studies. *Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association*, 2015; 13(8):1405–15.e1; quiz e130.

[496] Plöger S, Stumpff F, Penner GB, *et al.* Microbial butyrate and its role for barrier function in the gastrointestinal tract. *Annals of the New York Academy of Sciences*, 2012;1258:52–9.

[497] Poirier R, Jacquot S, Vaillend C, *et al.* Deletion of the Coffin-Lowry syndrome gene Rsk2 in mice is associated with impaired spatial learning and reduced control of exploratory behavior. *Behavior Genetics*, 2007;37(1):31–50.

[498] Polito JM, Childs B, Mellits ED, *et al.* Crohn's disease: influence of age at diagnosis on site and clinical type of disease. *Gastroenterology*, 1996;111:580–586.

[499] Polytarchou C, Hommes DW, Palumbo T, *et al.* MicroRNA214 Is Associated With Progression of Ulcerative Colitis, and Inhibition Reduces Development of Colitis and Colitis-Associated Cancer in Mice. *Gastroenterology*, 2015;149(4):981–992.

[500] Ponder A and Long MD. A clinical review of recent findings in the epidemiology of inflammatory bowel disease. *Clinical epidemiology*, 2013;5:237–47.

[501] Praper T, Sonnen A, Viero G, *et al.* Human perforin employs different avenues to damage membranes. *Journal of Biological Chemistry*, 2011;286(4):2946–2955.

[502] Pritchard CC, Kroh E, Wood B, *et al.* Blood cell origin of circulating microRNAs: a cautionary note for cancer biomarker studies. *Cancer prevention research*, 2012; 5(3):492–7.

[503] Probert CS, Jayanthi V, Pinder D, *et al.* Epidemiological study of ulcerative proctocolitis in Indian migrants and the indigenous population of Leicestershire. *Gut*, 1992;33(5):687–93.

[504] Probst AV, Dunleavy E, and Almouzni G. Epigenetic inheritance during the cell cycle. *Nature reviews. Molecular cell biology*, 2009;10(3):192–206.

[505] Punga AR, Andersson M, Alimohammadi M, *et al.* Disease specific signature of circulating miR-150-5p and miR-21-5p in myasthenia gravis patients. *Journal of the neurological sciences*, 2015;356(1-2):90–6.

[506] Radford-Smith GL, Edwards JE, Purdie DM, *et al.* Protective role of appendicectomy on onset and severity of ulcerative colitis and Crohn's disease. *Gut*, 2002;51(6):808–13.

[507] Raelson JV, Little RD, Ruether A, *et al.* Genome-wide association study for Crohn's disease in the Quebec Founder Population identifies multiple validated disease loci. *Proceedings of the National Academy of Sciences of the United States of America*, 2007;104(37):14747–52.

[508] Raine T, Liu JZ, Anderson CA, *et al.* Generation of primary human intestinal T cell transcriptomes reveals differential expression at genetic risk loci for immune-mediated disease. *Gut*, 2014;pp. 250–259.

[509] Rakyan VK, Down Ta, Balding DJ, *et al.* Epigenome-wide association studies for common human diseases. *Nature reviews. Genetics*, 2011;12(8):529–541.

[510] Ram PA and Waxman DJ. SOCS/CIS protein inhibition of growth hormone-stimulated STAT5 signaling by multiple mechanisms. *Journal of Biological Chemistry*, 1999;274(50):35553–35561.

[511] Randall-Demllo S, Chieppa M, and Eri R. Intestinal epithelium and autophagy: Partners in gut homeostasis. *Frontiers in Immunology*, 2013;4(SEP):1–14.

[512] Rath HC, Herfarth HH, Ikeda JS, *et al.* Normal luminal bacteria, especially Bacteroides species, mediate chronic colitis, gastritis, and arthritis in HLA-B27/human beta2 microglobulin transgenic rats. *The Journal of clinical investigation*, 1996;98(4):945–53.

[513] Regl G, Kasper M, Schnidar H, *et al.* Activation of the BCL2 promoter in response to Hedgehog/GLI signal transduction is predominantly mediated by GLI2. *Cancer research*, 2004;64(21):7724–31.

[514] Rennick DM, Fort MM, and Davidson NJ. Studies with IL-10-/- mice: an overview. *Journal of leukocyte biology*, 1997;61(4):389–396.

[515] Resta N, Giorda R, Bagnulo R, *et al.* Breakpoint determination of 15 large deletions in Peutz-Jeghers subjects. *Human Genetics*, 2010;128:373–382.

[516] Ribas J and Lupold SE. The transcriptional regulation of miR-21, its multiple transcripts, and their implication in prostate cancer. *Cell Cycle*, 2010;9(5):923–929.

[517] Ribas J, Ni X, Castanares M, *et al.* A novel source for miR-21 expression through the alternative polyadenylation of VMP1 gene transcripts. *Nucleic acids research*, 2012;40(14):6821–33.

[518] Richard-Miceli C and Criswell LA. Emerging patterns of genetic overlap across autoimmune disorders. *Genome Medicine*, 2012;4(1):6.

[519] Rigby RA and Stasinopoulos DM. Generalized additive models for location, scale and shape,(with discussion). *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 2005;54(3):507–554.

[520] Rigby RJ, Simmons JG, Greenhalgh CJ, *et al.* Suppressor of cytokine signaling 3 (SOCS3) limits damage-induced crypt hyper-proliferation and inflammation-associated tumorigenesis in the colon. *Oncogene*, 2007;26(33):4833–41.

[521] Rioux JD, Xavier RJ, Taylor KD, *et al.* Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. *Nature genetics*, 2007;39(5):596–604.

[522] Ripke S, O'Dushlaine C, Chambert K, *et al.* Genome-wide association analysis identifies 13 new risk loci for schizophrenia - SuppMat. *Nature Genetics*, 2013; 45(10):1150–9.

[523] Robertson DJ and Sandler RS. Measles Virus and Crohn ' s Disease : A Critical Appraisal of the Current Literature. *Inflammatory Bowel Disease*, 2001;7(1):51–57.

[524] Rogler G. The history and philosophy of inflammatory bowel disease. *Digestive diseases (Basel, Switzerland)*, 2013;31(3-4):270–7.

[525] Rolhion N, Barnich N, Bringer MA, *et al.* Abnormally expressed ER stress response chaperone Gp96 in CD favours adherent-invasive Escherichia coli invasion. *Gut*, 2010;59(10):1355–1362.

[526] Romeo Y, Zhang X, and Roux PP. Regulation and function of the RSK family of protein kinases. *The Biochemical journal*, 2012;441(2):553–69.

[527] Ropolo A, Grasso D, Pardo R, *et al.* The pancreatitis-induced vacuole membrane protein 1 triggers autophagy in mammalian cells. *The Journal of biological chemistry*, 2007;282(51):37124–33.

[528] Rose CD, Doyle TM, McIlvain-Simpson G, *et al.* Blau syndrome mutation of CARD15/NOD2 in sporadic early onset granulomatous arthritis. *J Rheumatol.*, 2005;32(0315-162X):373–375.

[529] Rothfuss KS, Stange EF, and Herrlinger KR. Extraintestinal manifestations and complications in inflammatory bowel diseases. *World Journal of Gastroenterology*, 2006;12(30):4819–4831.

[530] Rottenberg ME and Carow B. SOCS3 and STAT3, major controllers of the outcome of infection with Mycobacterium tuberculosis. *Seminars in immunology*, 2014;26(6):518–532.

[531] Ruemmele FM, El Khoury MG, Talbotec C, *et al.* Characteristics of inflammatory bowel disease with onset during the first year of life. *Journal of pediatric gastroenterology and nutrition*, 2006;43(5):603–609.

[532] Rufini S, Ciccacci C, Di Fusco D, *et al.* Autophagy and inflammatory bowel disease: Association between variants of the autophagy-related IRGM gene and susceptibility to Crohn's disease. *Digestive and Liver Disease*, 2015;47(9):744–750.

[533] Ruiz I Altaba A. Gli proteins encode context-dependent positive and negative functions: implications for development and disease. *Development (Cambridge, England)*, 1999;126(14):3205–3216.

[534] Ruiz-Lafuente N, Alcaraz-García MJ, Sebastián-Ruiz S, *et al.* The Gene Expression Response of Chronic Lymphocytic Leukemia Cells to IL-4 Is Specific, Depends on ZAP-70 Status and Is Differentially Affected by an NFκB Inhibitor. *PloS one*, 2014;9(10):e109533.

[535] Ruiz-Lafuente N, Alcaraz-García MJ, Sebastián-Ruiz S, *et al.* IL-4 Up-Regulates MiR-21 and the MiRNAs Hosted in the CLCN5 Gene in Chronic Lymphocytic Leukemia. *PloS one*, 2015;10(4):e0124936.

[536] Runtsch MC, Round JL, and O'Connell RM. MicroRNAs and the regulation of intestinal homeostasis. *Frontiers in genetics*, 2014;5(347).

[537] Rutgeerts P, Goboes K, Peeters M, *et al.* Effect of faecal stream diversion on recurrence of Crohn's disease in the neoterminal ileum. *Lancet*, 1991;338(8770):771–774.

[538] Rutz S and Ouyang W. Regulation of interleukin-10 and interleukin-22 expression in T helper cells. *Current Opinion in Immunology*, 2011;23(5):605–612.

[539] Saadati M and Benner A. Statistical challenges of high-dimensional methylation data. *Statistics in medicine*, 2014;33(30):5347–57.

[540] Saccani S and Natoli G. Dynamic changes in histone H3 Lys 9 methylation occurring at tightly regulated inducible inflammatory genes. *Genes & development*, 2002;16(17):2219–24.

[541] Saez-Lara MJ, Gomez-Llorente C, Plaza-Diaz J, *et al.* The Role of Probiotic Lactic Acid Bacteria and Bifidobacteria in the Prevention and Treatment of Inflammatory Bowel Disease and Other Related Diseases: A Systematic Review of Randomized Human Clinical Trials. *BioMed Research International*, 2015; 2015:1–15.

[542] Saito S, Kato J, Hiraoka S, *et al.* DNA methylation of colon mucosa in ulcerative colitis patients: Correlation with inflammatory status. *Inflammatory Bowel Diseases*, 2011;17(9):1955–1965.

[543] Saito Y, Liang G, Egger G, *et al.* Specific activation of microRNA-127 with downregulation of the proto-oncogene BCL6 by chromatin-modifying drugs in human cancer cells. *Cancer cell*, 2006;9(6):435–43.

[544] Saito-Ohara F, Imoto I, Inoue J, *et al.* PPM1D is a potential target for 17q gain in neuroblastoma. *Cancer research*, 2003;63(8):1876–83.

[545] Salzberg A, Fisher O, Siman-Tov R, *et al.* Identification of methylated sequences in genomic DNA of adult Drosophila melanogaster. *Biochemical and biophysical research communications*, 2004;322(2):465–9.

[546] Salzman J, Gawad C, Wang PL, *et al.* Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. *PloS one*, 2012;7(2):e30733.

[547] Sato F, Harpaz N, Shibata D, *et al.* Hypermethylation of the p14ARF gene in ulcerative colitis-associated colorectal carcinogenesis. *Cancer Research*, 2002; 62(4):1148–1151.

[548] Sato F, Shibata D, Harpaz N, *et al.* Aberrant methylation of the HPP1 gene in ulcerative colitis-associated colorectal carcinoma. *Cancer Research*, 2002; 62(23):6820–6822.

[549] Sauermann M, Sahin O, Sültmann H, *et al.* Reduced expression of vacuole membrane protein 1 affects the invasion capacity of tumor cells. *Oncogene*, 2008; 27(9):1320–6.

[550] Saunders LR and Verdin E. Sirtuins: critical regulators at the crossroads between cancer and aging. *Oncogene*, 2007;26(37):5489–504.

[551] Sauntharajah Y, Sekeres M, Advani A, *et al.* Evaluation of noncytotoxic DNMT1-depleting therapy in patients with myelodysplastic syndromes. *The Journal of clinical investigation*, 2015;125(3):1043–1055.

[552] Sawant DV, Wu H, Kaplan MH, *et al.* The Bcl6 target gene microRNA-21 promotes Th2 differentiation by a T cell intrinsic pathway. *Molecular immunology*, 2013;54(3-4):435–442.

[553] Scharl M, Mwinyi J, Fischbeck A, *et al.* Crohn's disease-associated polymorphism within the PTPN2 gene affects muramyl-dipeptide-induced cytokine secretion and autophagy. *Inflammatory bowel diseases*, 2012;18(5):900–12.

[554] Scharl M, Wojtal KA, Becker HM, *et al.* Protein tyrosine phosphatase non-receptor type 2 regulates autophagosome formation in human intestinal cells. *Inflammatory bowel diseases*, 2012;18(7):1287–302.

[555] Schilderink R, Verseijden C, and de Jonge WJ. Dietary inhibitors of histone deacetylases in intestinal immunity and homeostasis. *Frontiers in immunology*, 2013;4:226.

[556] Schmitz M, Driesch C, Jansen L, *et al.* Non-random integration of the HPV genome in cervical cancer. *PloS one*, 2012;7(6):e39632.

[557] Schreiber S, Fedorak RN, Nielsen OH, *et al.* Safety and efficacy of recombinant human interleukin 10 in chronic active Crohn's disease. Crohn's Disease IL-10 Cooperative Study Group. *Gastroenterology*, 2000;119(6):1461–1472.

[558] Schübeler D. Epigenetic Islands in a Genetic Ocean. *Science*, 2012;338(6108):756–757.

[559] Sebastian S, Hernández V, Myrelid P, *et al.* Colorectal cancer in inflammatory bowel disease: Results of the 3rd ECCO pathogenesis scientific workshop (I). *Journal of Crohn's and Colitis*, 2014;8(1):5–18.

[560] Segain JP, Raingeard de la Blétière D, Bourreille A, *et al.* Butyrate inhibits inflammatory responses through NFkappaB inhibition: implications for Crohn's disease. *Gut*, 2000;47(3):397–403.

[561] Sellon RK, Tonkonogy S, Schultz M, *et al.* Resident enteric bacteria are necessary for development of spontaneous colitis and immune system activation in interleukin-10-deficient mice. *Infection and immunity*, 1998;66(11):5224–31.

[562] Setoguchi R, Tachibana M, Naoe Y, *et al.* Repression of the transcription factor Th-POK by Runx complexes in cytotoxic T cell development. *Science (New York, N.Y.)*, 2008;319(5864):822–825.

[563] Shabalin AA. Matrix eQTL: Ultra fast eQTL analysis via large matrix operations. *Bioinformatics*, 2012;28(10):1353–1358.

[564] Shaw SY, Blanchard JF, and Bernstein CN. Early childhood measles vaccinations are not associated with paediatric IBD: a population-based analysis. *Journal of Crohn's & colitis*, 2015;9(4):334–8.

[565] Shi C, Liang Y, Yang J, *et al.* MicroRNA-21 knockout improve the survival rate in DSS induced fatal colitis through protecting against inflammation and tissue injury. *PloS one*, 2013;8(6):e66814.

[566] Shim JO, Hwang S, Yang HR, *et al.* Interleukin-10 receptor mutations in children with neonatal-onset Crohn's disease and intractable ulcerating enterocolitis. *European journal of gastroenterology & hepatology*, 2013;25(10):1235–40.

[567] Shim JO and Seo JK. Very early-onset inflammatory bowel disease (IBD) in infancy is a different disease entity from adult-onset IBD; one form of interleukin-10 receptor mutations. *Journal of human genetics*, 2014;59(6):337–41.

[568] Shimo T, Tachibana K, Saito K, *et al.* Design and evaluation of locked nucleic acid-based splice-switching oligonucleotides in vitro. *Nucleic Acids Research*, 2014;42(12):8174–8187.

[569] Shukla S, Kavak E, Gregory M, *et al.* CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature*, 2011;479(7371):74–9.

[570] Šidák Z. Rectangular Confidence Regions for the Means of Multivariate Normal Distributions. *Journal of the American Statistical Association*, 1967;62(318):626–33.

[571] Sikora SK, Spady D, Prosser C, *et al.* Trace elements and vitamins at diagnosis in pediatric-onset inflammatory bowel disease. *Clinical Pediatrics*, 2011;50(6):488–492.

[572] Silverberg MS, Cho JH, Rioux JD, *et al.* Ulcerative colitis-risk loci on chromosomes 1p36 and 12q15 found by genome-wide association study. *Nature genetics*, 2009;41(2):216–220.

[573] Silverberg MS, Satsangi J, Ahmad T, *et al.* Toward an integrated clinical, molecular and serological classification of inflammatory bowel disease: report of a Working Party of the 2005 Montreal World Congress of Gastroenterology. *Canadian journal of gastroenterology*, 2005;19 Suppl A:5A–36A.

[574] Simas AB, Barreto-Souza W, and Rocha AV. Improved estimators for a general class of beta regression models. *Computational Statistics and Data Analysis*, 2010;54(2):348–366.

[575] Singh N, Gurav A, Sivaprakasam S, *et al.* Activation of Gpr109a, receptor for niacin and the commensal metabolite butyrate, suppresses colonic inflammation and carcinogenesis. *Immunity*, 2014;40(1):128–39.

[576] Singh S, Kullo IJ, Pardi DS, *et al.* Epidemiology, risk factors and management of cardiovascular diseases in IBD. *Nature reviews. Gastroenterology & hepatology*, 2015;12(1):26–35.

[577] Singhal R, Taylor J, Owoniyi M, *et al.* The role of appendicectomy in the subsequent development of inflammatory bowel disease: A UK-based study. *International Journal of Colorectal Disease*, 2010;25(4):509–513.

[578] Skroza N, Proietti I, Pampena R, *et al.* Correlations between psoriasis and inflammatory bowel diseases. *BioMed research international*, 2013;2013:983902.

[579] Slattery ML, Lundgreen A, Herrick JS, *et al.* Diet and colorectal cancer: analysis of a candidate pathway using SNPS, haplotypes, and multi-gene assessment. *Nutrition and cancer*, 2011;63(8):1226–34.

[580] Slattery ML, Lundgreen A, Herrick JS, *et al.* Genetic variation in RPS6KA1, RPS6KA2, RPS6KB1, RPS6KB2, and PDK1 and risk of colon or rectal cancer. *Mutation research*, 2011;706(1-2):13–20.

[581] Smith AM, Rahman FZ, Hayee B, *et al.* Disordered macrophage cytokine secretion underlies impaired acute inflammation and bacterial clearance in Crohn's disease. *The Journal of experimental medicine*, 2009;206(9):1883–1897.

[582] Smyth GK. Limma: linear models for microarray data. In Gentleman R, Carey V, Dudoit S, *et al.*, eds., *Bioinformatics and Computational Biology Solutions Using {R} and Bioconductor*, pp. 397–420. Springer, New York, 2005;.

[583] So J, Pasculescu A, Dai AY, *et al.* Integrative analysis of kinase networks in TRAIL-induced apoptosis provides a source of potential targets for combination therapy. *Sci. Signal*, 2015;8(371):13–15.

[584] Sokol H, Pigneur B, Watterlot L, *et al.* Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proceedings of the National Academy of Sciences of the United States of America*, 2008;105(43):16731–6.

[585] Song MS and Rossi JJ. The anti-miR21 antagomir, a therapeutic tool for colorectal cancer, has a potential synergistic effect by perturbing an angiogenesis-associated miR30. *Frontiers in genetics*, 2014;4(301).

[586] Soon IS, DeBruyn JCC, and Wrobel I. Immunization history of children with inflammatory bowel disease. *Canadian journal of gastroenterology*, 2013;27(4):213–6.

[587] Soon IS, Molodecky NA, Rabi DM, *et al.* The relationship between urban environment and the inflammatory bowel diseases: a systematic review and meta-analysis. *BMC gastroenterology*, 2012;12(1):51.

[588] Stadler MB, Murr R, Burger L, *et al.* DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature*, 2011;480(7378):490–5.

[589] Steenholdt C, Andresen L, Pedersen G, *et al.* Expression and function of toll-like receptor 8 and Tollip in colonic epithelial cells from patients with inflammatory bowel disease. *Scandinavian journal of gastroenterology*, 2009;44(2):195–204.

[590] Stevens C, Henderson P, Nimmo ER, *et al.* The intermediate filament protein, vimentin, is a regulator of NOD2 activity. *Gut*, 2013;62(5):695–707.

[591] Strachan DP. Hay fever, hygiene, and household size. *BMJ (Clinical research ed.)*, 1989;299(6710):1259–1260.

[592] Strahl BD and Allis CD. The language of covalent histone modifications. *Nature*, 2000;403(6765):41–5.

[593] Sulzmaier FJ and Ramos JW. RSK isoforms in cancer cell invasion and metastasis. *Cancer Research*, 2013;73(20):6099–6105.

[594] Suzuki H, Yamamoto E, Maruyama R, *et al.* Biological significance of the CpG island methylator phenotype. *Biochemical and Biophysical Research Communications*, 2014;455(1-2):35–42.

[595] Szkudlapski D, Labuzek K, Pokora Z, *et al.* The emering role of helminths in treatment of the inflammatory bowel disorders. *Journal of physiology and pharmacology : an official journal of the Polish Physiological Society*, 2014;65(6):741–51.

[596] Tahara T, Shibata T, Nakamura M, *et al.* Effect of MDR1 gene promoter methylation in patients with ulcerative colitis. *International Journal of Molecular Medicine*, 2009;23(4):521–527.

[597] Tahara T, Shibata T, Nakamura M, *et al.* Promoter methylation of protease-activated receptor (PAR2) is associated with severe clinical phenotypes of ulcerative colitis (UC). *Clinical and Experimental Medicine*, 2009;9(2):125–130.

[598] Tahara T, Shibata T, Nakamura M, *et al.* Association between polymorphisms in the XRCC1 and GST genes, and CpG island methylation status in colonic mucosa in ulcerative colitis. *Virchows Archiv : an international journal of pathology*, 2011; 458(2):205–11.

[599] Tahara T, Shibata T, Nakamura M, *et al.* Host genetic factors, related to inflammatory response, influence the CpG island methylation status in colonic mucosa in ulcerative colitis. *Anticancer research*, 2011;31(3):933–8.

[600] Takagi T, Naito Y, Mizushima K, *et al.* Increased expression of microRNA in the inflamed colonic mucosa of patients with active ulcerative colitis. *Journal of gastroenterology and hepatology*, 2010;25(Suppl 1):S129–33.

[601] Takahashi K, Sugi Y, Hosono A, *et al.* Epigenetic regulation of TLR4 gene expression in intestinal epithelial cells for the maintenance of intestinal homeostasis. *Journal of immunology*, 2009;183(10):6522–9.

[602] Takano A, Zochi R, and Hibi M. Function of strawberry notch Family Genes in the Zebrafish Brain Development. *The Kobe Journal of Medical Science*, 2010; 56(5):220–230.

[603] Takano A, Zochi R, Hibi M, *et al.* Expression of strawberry notch family genes during zebrafish embryogenesis. *Developmental dynamics : an official publication of the American Association of Anatomists*, 2010;239(6):1789–96.

[604] Takeshima H, Ikegami D, Wakabayashi M, *et al.* Induction of aberrant trimethylation of histone H3 lysine 27 by inflammation in mouse colonic epithelial cells. *Carcinogenesis*, 2012;33(12):2384–90.

[605] Takeshima H, Niwa T, Takahashi T, *et al.* Frequent involvement of chromatin remodeler alterations in gastric field cancerization. *Cancer Letters*, 2015; 357(1):328–338.

[606] Tan G, Zeng B, and Zhi FC. Regulation of human enteric α-defensins by NOD2 in the Paneth cell lineage. *European journal of cell biology*, 2015;94(1):60–6.

[607] Tanabe T, Chamaillard M, Ogura Y, *et al.* Regulatory regions and critical residues of NOD2 involved in muramyl dipeptide recognition. *Embo J*, 2004; 23(7):1587–1597.

[608] Taniuchi I, Osato M, Egawa T, *et al.* Differential requirements for Runx proteins in CD4 repression and epigenetic silencing during T lymphocyte development. *Cell*, 2002;111(5):621–33.

[609] Tannahill GM, Elliott J, Barry AC, *et al.* SOCS2 Can Enhance Interleukin-2 (IL-2) and IL-3 Signaling by Accelerating SOCS3 Degradation. *Molecular and Cellular Biology*, 2005;25(20):9115–9126.

[610] Tay FC, Lim JK, Zhu H, *et al.* Using artificial microRNA sponges to achieve microRNA loss-of-function in cancer cells. *Advanced drug delivery reviews*, 2015; 81:117–27.

[611] Tenenboim H, Smirnova J, Willmitzer L, *et al.* VMP1-deficient Chlamydomonas exhibits severely aberrant cell morphology and disrupted cytokinesis. *BMC plant biology*, 2014;14(1):121.

[612] Teschendorff AE, Marabita F, Lechner M, *et al.* A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics*, 2013;29(2):189–196.

[613] Teslovich TM, Musunuru K, Smith AV, *et al.* Biological, clinical and population relevance of 95 loci for blood lipids. *Nature*, 2010;466(7307):707–713.

[614] Thagia I, Shaw EJ, Smith E, *et al.* Intestinal epithelial suppressor of cytokine signaling 3 enhances microbial-induced inflammatory tumor necrosis factor-α, contributing to epithelial barrier dysfunction. *American journal of physiology. Gastrointestinal and liver physiology*, 2015;308(1):G25–31.

[615] Thayer WR, Coutu JA, Chiodini RJ, *et al.* Possible role of mycobacteria in inflammatory bowel disease. II. Mycobacterial antibodies in Crohn's disease. *Digestive diseases and sciences*, 1984;29(12):1080–5.

[616] The Bioconductor Dev Team. BSgenome.Hsapiens.UCSC.gs19: Full genome sequences for Homo sapiens (UCSC version hg19).

[617] The IBD International Genetics Consortium. International collaboration provides convincing linkage replication in complex disease through analysis of a large pooled data set: Crohn disease and chromosome 16. *American journal of human genetics*, 2001;68(5):1165–71.

[618] Thibault R, Blachier F, Darcy-Vrillon B, *et al.* Butyrate utilization by the colonic mucosa in inflammatory bowel diseases: a transport deficiency. *Inflammatory bowel diseases*, 2010;16(4):684–95.

[619] Thjodleifsson B, Geirsson AJ, Björnsson S, *et al.* A common genetic background for inflammatory bowel disease and ankylosing spondylitis: a genealogic study in Iceland. *Arthritis and rheumatism*, 2007;56(8):2633–2639.

[620] Thompson N, Montgomery S, Pounder R, *et al.* Is measles vaccination a risk factor for inflammatory bowel disease? *Lancet*, 1995;345(8957):1071–1074.

[621] Tian Y, Li Z, Hu W, *et al.* C. elegans screen identifies autophagy genes specific to multicellular organisms. *Cell*, 2010;141(6):1042–55.

[622] Tillman LG, Geary RS, and Hardee GE. Oral delivery of antisense oligonucleotides in man. *Journal of pharmaceutical sciences*, 2008;97(1):225–36.

[623] Ting JPY and Davis BK. CATERPILLER: a novel gene family important in immunity, cell death, and diseases. *Annual review of immunology*, 2005;23:387–414.

[624] Ting JPY, Lovering RC, Alnemri ES, *et al.* The NLR Gene Family: A Standard Nomenclature. *Immunity*, 2008;28(3):285–287.

[625] Tokuhira N, Kitagishi Y, Suzuki M, *et al.* PI3K/AKT/PTEN pathway as a target for Crohn's disease therapy (Review). *International Journal of Molecular Medicine*, 2015;35(1):10–16.

[626] Tokuhiro S, Yamada R, Chang X, *et al.* An intronic SNP in a RUNX1 binding site of SLC22A4, encoding an organic cation transporter, is associated with rheumatoid arthritis. *Nature Genetics*, 2003;35(4):341–348.

[627] Tominaga K, Fujii S, Mukawa K, *et al.* Prediction of colorectal neoplasia by quantitative methylation analysis of estrogen receptor gene in nonneoplastic epithelium from patients with ulcerative colitis. *Clinical cancer research*, 2005; 11(24):8880–5.

[628] Törkvist L, Halfvarson J, Ong RTH, *et al.* Analysis of 39 Crohn's disease risk loci in Swedish inflammatory bowel disease patients. *Inflammatory Bowel Diseases*, 2010;16(6):907–909.

[629] Tornatore L, Thotakura AK, Bennett J, *et al.* The nuclear factor kappa B signaling pathway: Integrating metabolism with inflammation. *Trends in Cell Biology*, 2012;22(11):557–566.

[630] Trabzuni D, Ryten M, Emmett W, *et al.* Fine-mapping, gene expression and splicing analysis of the disease associated LRRK2 locus. *PloS one*, 2013;8(8):e70724.

[631] Travassos LH, Carneiro LAM, Ramjeet M, *et al.* Nod1 and Nod2 direct autophagy by recruiting ATG16L1 to the plasma membrane at the site of bacterial entry. *Nature immunology*, 2010;11(1):55–62.

[632] Tremaine WJ. Is Indeterminate Colitis Determinable? *Current Gastroenterology Reports*, 2012;14(2):162–165.

[633] Trerotola M, Relli V, Simeone P, *et al.* Epigenetic inheritance and the missing heritability. *Human Genomics*, 2015;9(1):17.

[634] Trivier E, De Cesare D, Jacquot S, *et al.* Mutations in the kinase Rsk-2 associated with Coffin-Lowry syndrome. *Nature*, 1996;384(6609):567–570.

[635] Tsaprouni LG, Ito K, Powell JJ, *et al.* Differential patterns of histone acetylation in inflammatory bowel diseases. *Journal of inflammation*, 2011;8(1):1.

[636] Tsaprouni LG, Yang T, Bell J, *et al.* Epigenetics Cigarette smoking reduces DNA methylation levels at multiple genomic loci but the effect is partially reversible upon cessation. *Epigenetics*, 2014;9(10):1382–1396.

[637] Tsuboi K, Nishitani M, Takakura A, *et al.* Autophagy protects against colitis by the maintenance of normal gut microflora and secretion of mucus. *Journal of Biological Chemistry*, 2015;290(33):20511–20526.

[638] Turgeon N, Blais M, Gagné JM, *et al.* HDAC1 and HDAC2 restrain the intestinal inflammatory response by regulating intestinal epithelial cell differentiation. *PloS one*, 2013;8(9):e73785.

[639] Turgeon N, Gagné JM, Blais M, *et al.* The acetylome regulators Hdac1 and Hdac2 differently modulate intestinal epithelial cell dependent homeostatic responses in experimental colitis. *American journal of physiology. Gastrointestinal and liver physiology*, 2014;306(7):G594–605.

[640] Turnbaugh PJ, Quince C, Faith JJ, *et al.* Organismal, genetic, and transcriptional variation in the deeply sequenced gut microbiomes of identical twins. *Proceedings of the National Academy of Sciences*, 2010;107(16):7503–7508.

[641] Uhlig HH, Schwerd T, Koletzko S, *et al.* The diagnostic approach to monogenic very early onset inflammatory bowel disease. *Gastroenterology*, 2014;147(5):990–1007.

[642] UK IBD Genetics Consortium and Wellcome Trust Case Control Consortium. Genome-wide association study of ulcerative colitis identifies three new susceptibility loci, including the HNF4A region. *Nature genetics*, 2009;41(12):1330–1334.

[643] Umeno J, Asano K, Matsushita T, *et al.* Meta-analysis of published studies identified eight additional common susceptibility loci for Crohn's disease and ulcerative colitis. *Inflammatory Bowel Diseases*, 2011;17(12):2407–2415.

[644] Ushijima T. Epigenetic field for cancerization. *Journal of biochemistry and molecular biology*, 2007;40(2):142–150.

[645] Uzel G, Kleiner DE, Kuhns DB, *et al.* Dysfunctional LAD-1 Neutrophils and Colitis. *Gastroenterology*, 2001;121(4):958–964.

[646] Uzel G, Tng E, Rosenzweig SD, *et al.* Reversion mutations in patients with leukocyte adhesion deficiency type-1 (LAD-1). *Blood*, 2007;111(1):209–218.

[647] Vaccaro MI, Grasso D, Ropolo A, *et al.* VMP1 expression correlates with acinar cell cytoplasmic vacuolization in arginine-induced acute pancreatitis. *Pancreatology*, 2003;3(1):69–74.

[648] Vaccaro MI, Ropolo A, Grasso D, *et al.* A novel mammalian trans-membrane protein reveals an alternative initiation pathway for autophagy. *Autophagy*, 2008;4(3):388–90.

[649] Valen E, Preker P, Andersen PR, *et al.* Biogenic mechanisms and utilization of small RNAs derived from human protein-coding genes. *Nature structural & molecular biology*, 2011;18(9):1075–82.

[650] van Deventer SJ, Elson CO, and Fedorak RN. Multiple doses of intravenous interleukin 10 in steroid-refractory Crohn's disease. Crohn's Disease Study Group. *Gastroenterology*, 1997;113(2):383–389.

[651] Van Guelpen B, Hultdin J, Johansson I, *et al.* Low folate levels may protect against colorectal cancer. *Gut*, 2006;55(10):1461–6.

[652] van Heel DA, Fisher SA, Kirby A, *et al.* Inflammatory bowel disease susceptibility loci defined by genome scan meta-analysis of 1952 affected relative pairs. *Human Molecular Genetics*, 2004;13(7):763–770.

[653] Van Limbergen J, Russell RK, Drummond HE, *et al.* Definition of phenotypic characteristics of childhood-onset inflammatory bowel disease. *Gastroenterology*, 2008;135(4):1114–22.

[654] Van Limbergen J, Russell RK, Nimmo ER, *et al.* Autophagy gene ATG16L1 influences susceptibility and disease location but not childhood-onset in Crohn's disease in northern Europe. *Inflammatory Bowel Diseases*, 2008;14(3):338–346.

[655] Van Neste L, Herman JG, Otto G, *et al.* The epigenetic promise for prostate cancer diagnosis. *The Prostate*, 2012;72(11):1248–61.

[656] van Nood E, Vrieze A, Nieuwdorp M, *et al.* Duodenal Infusion of Donor Feces for Recurrent Clostridium difficile. *New England Journal of Medicine*, 2013; 368(5):407–415.

[657] Vandesompele J, De Preter K, Pattyn F, *et al.* Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome biology*, 2002;3(7):RESEARCH0034.

[658] Vastert SJ, van Wijk R, D'Urbano LE, *et al.* Mutations in the perforin gene can be linked to macrophage activation syndrome in patients with systemic onset juvenile idiopathic arthritis. *Rheumatology (Oxford, England)*, 2010;49(3):441–449.

[659] Vaughn BP. Prevention of post-operative recurrence of Crohn's disease. *World Journal of Gastroenterology*, 2014;20(5):1147.

[660] Vavricka SR and Rogler G. Intestinal absorption and vitamin levels: is a new focus needed? *Digestive diseases (Basel, Switzerland)*, 2012;30(suppl 3):73–80.

[661] Vavricka SR, Schoepfer A, Scharl M, *et al.* Extraintestinal Manifestations of Inflammatory Bowel Disease. *Inflammatory bowel diseases*, 2015;21(8):1982–92.

[662] Venables WN and Ripley BD. *Modern Applied Statistics with S*. Springer, New York, fourth ed., 2002.

[663] Ventham NT, Kennedy NA, Adams AT, *et al.* Integrative epigenome-wide analysis demonstrates that DNA methylation may mediate genetic risk in inflammatory bowel disease. *Nature communications*, 2016;7:13507.

[664] Ventham NT, Kennedy NA, Duffy A, *et al.* Comparison of mortality following hospitalisation for ulcerative colitis in Scotland between 1998-2000 and 2007-2009. *Alimentary pharmacology & therapeutics*, 2014;39(12):1387–97.

[665] Vermeire S, Rutgeerts P, Van Steen K, *et al.* Genome wide scan in a Flemish inflammatory bowel disease population: support for the IBD4 locus, population heterogeneity, and epistasis. *Gut*, 2004;53(7):980–986.

[666] Viala J, Chaput C, Boneca IG, *et al.* Nod1 responds to peptidoglycan delivered by the Helicobacter pylori cag pathogenicity island. *Nature immunology*, 2004; 5(11):1166–1174.

[667] Vieira ELM, Leonel AJ, Sad AP, *et al.* Oral administration of sodium butyrate attenuates inflammation and mucosal lesion in experimental acute ulcerative colitis. *The Journal of nutritional biochemistry*, 2012;23(5):430–6.

[668] Villani AC, Lemire M, Fortin G, *et al.* Common variants in the NLRP3 region contribute to Crohn's disease susceptibility. *Nature genetics*, 2009;41(1):71–6.

[669] Visscher PM, Brown MA, McCarthy MI, *et al.* Five years of GWAS discovery. *American journal of human genetics*, 2012;90(1):7–24.

[670] Voegtlin M, Vavricka SR, Schoepfer AM, *et al.* Prevalence of anaemia in inflammatory bowel disease in Switzerland: a cross-sectional study in patients from private practices and university hospitals. *Journal of Crohn's & colitis*, 2010; 4(6):642–8.

[671] Vojinovic J, Damjanov N, D'Urzo C, *et al.* Safety and efficacy of an oral histone deacetylase inhibitor in systemic-onset juvenile idiopathic arthritis. *Arthritis and rheumatism*, 2011;63(5):1452–8.

[672] von Kampen O, Lipinski S, Till A, *et al.* Caspase recruitment domain-containing protein 8 (CARD8) negatively regulates NOD2-mediated signaling. *The Journal of biological chemistry*, 2010;285(26):19921–6.

[673] Waetzig GH, Seegert D, Rosenstiel P, *et al.* p38 mitogen-activated protein kinase is activated and linked to TNF-alpha signaling in inflammatory bowel disease. *Journal of immunology*, 2002;168(10):5342–51.

[674] Wagner RN, Proell M, Kufer TA, *et al.* Evaluation of Nod-like receptor (NLR) effector domain interactions. *PloS one*, 2009;4(4):e4931.

[675] Wahl S, Fenske N, Zeilinger S, *et al.* On the potential of models for location and scale for genome-wide DNA methylation data. *BMC Bioinformatics*, 2014; 15(1):232.

[676] Wakefield AJ, Ekbom A, Dhillon AP, *et al.* Crohn's disease: pathogenesis and persistent measles virus infection. *Gastroenterology*, 1995;108(3):911–6.

[677] Wakefield AJ, Pittilo RM, Sim R, *et al.* Evidence of persistent measles virus infection in Crohn's disease. *Journal of medical virology*, 1993;39(4):345–53.

[678] Wallings R, Manzoni C, and Bandopadhyay R. Cellular processes associated with LRRK2 function and dysfunction. *FEBS Journal*, 2015;282(15):2806–2826.

[679] Walsh CP and Bestor TH. Cytosine methylation and mammalian development. *Genes & development*, 1999;13(1):26–34.

[680] Wang J, Zhou W, Cheng S, *et al.* Downregulation of Sprouty homology 2 by microRNA-21 inhibits proliferation, metastasis and invasion, however promotes the apoptosis of multiple myeloma cells. *Molecular Medicine Reports*, 2015;pp. 1810–1816.

[681] Wang T, Cai G, Qiu Y, *et al.* Structural segregation of gut microbiota between colorectal cancer patients and healthy volunteers. *The ISME journal*, 2012; 6(2):320–9.

[682] Wang X, Shaffer JR, Zeng Z, *et al.* Genome-wide association scan of dental caries in the permanent dentition. *BMC oral health*, 2012;12:57.

[683] Wang Y, Gao X, Wei F, *et al.* Diagnostic and prognostic value of circulating miR-21 for cancer: a systematic review and meta-analysis. *Gene*, 2014;533(1):389–97.

[684] Wang Z, Liu M, Nie X, *et al.* NOD1 and NOD2 control the invasiveness of trophoblast cells via the MAPK/p38 signaling pathway in human first-trimester pregnancy. *Placenta*, 2015;36(6):652–660.

[685] Warren JD, Xiong W, Bunker AM, *et al.* Septin 9 methylated DNA is a sensitive and specific blood test for colorectal cancer. *BMC Medicine*, 2011;9(1):133.

[686] Wasan SK, Coukos JA, and Farraye FA. Vaccinating the inflammatory bowel disease patient: Deficiencies in gastroenterologists knowledge. *Inflammatory Bowel Diseases*, 2011;17(12):2536–2540.

[687] Waterman M, Xu W, Stempak JM, *et al.* Distinct and overlapping genetic loci in crohn's disease and ulcerative colitis: Correlations with pathogenesis. *Inflammatory Bowel Diseases*, 2011;17(9):1936–1942.

[688] Wedlake L, Slack N, Andreyev HJN, *et al.* Fiber in the treatment and maintenance of inflammatory bowel disease: a systematic review of randomized controlled trials. *Inflammatory bowel diseases*, 2014;20(3):576–86.

[689] Weersma RK, Zhou L, Nolte IM, *et al.* Runt-related transcription factor 3 is associated with ulcerative colitis and shows epistasis with solute carrier family 22, members 4 and 5. *Inflammatory bowel diseases*, 2008;14(12):1615–22.

[690] Weinhold L, Wahl S, and Schmid M. A Statistical Model for the Analysis of Beta Values in DNA Methylation Studies. 2016;p. arXiv:1607.07028v1 [stat.ME].

[691] Weinstock JV and Elliott DE. Helminths and the IBD hygiene hypothesis. *Inflammatory Bowel Diseases*, 2009;15(1):128–133.

[692] Weinstock JV and Elliott DE. Translatability of helminth therapy in inflammatory bowel diseases. *International Journal for Parasitology*, 2013;43(3-4):245–251.

[693] Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, 2007; 447(7145):661–78.

[694] Wheeler JM, Kim HC, Efstathiou JA, *et al.* Hypermethylation of the promoter region of the E-cadherin gene (CDH1) in sporadic and ulcerative colitis associated colorectal cancer. *Gut*, 2001;48(3):367–371.

[695] Wilks S. Morbid appearances in the intestines of Miss Bankes. *Medical Times and Gazette*, 1859;19(September 10th):264.

[696] Wilks S and Moxon W. *Lectures on pathological anatomy.* Lindsay and Blakiston, Philadelphia, 2nd ed., 1875.

[697] Wittmann J and Jäck HM. Serum microRNAs as powerful cancer biomarkers. *Biochimica et biophysica acta*, 2010;1806(2):200–7.

[698] Wong PM, Puente C, Ganley IG, *et al.* The ULK1 complex: sensing nutrient signals for autophagy activation. *Autophagy*, 2013;9(2):124–37.

[699] Wong YF, Jakt LM, and Nishikawa SI. Prolonged treatment with DNMT inhibitors induces distinct effects in promoters and gene-bodies. *PloS one*, 2013; 8(8):e71099.

[700] Wouters CH, Maes A, Foley KP, *et al.* Blau Syndrome, the prototypic autoinflammatory granulomatous disease. *Pediatric rheumatology online journal*, 2014;12(1):33.

[701] Wu F, Dong F, Arendovich N, *et al.* Divergent influence of microRNA-21 deletion on murine colitis phenotypes. *Inflammatory Bowel Diseases*, 2014;20(11):1972–1985.

[702] Wu F, Guo NJ, Tian H, *et al.* Peripheral blood MicroRNAs distinguish active ulcerative colitis and Crohn's disease. *Inflammatory Bowel Diseases*, 2011; 17(1):241–250.

[703] Wu F, Zhang S, Dassopoulos T, *et al.* Identification of microRNAs associated with ileal and colonic Crohn's disease. *Inflammatory bowel diseases*, 2010;16(10):1729–38.

[704] Wu F, Zikusoka M, Trindade A, *et al.* MicroRNAs are differentially expressed in ulcerative colitis and alter expression of macrophage inflammatory peptide-2 alpha. *Gastroenterology*, 2008;135(5):1624–1635.e24.

[705] Wu H, Ng BSH, and Thibault G. Endoplasmic reticulum stress response in yeast and humans. *Bioscience reports*, 2014;34(4):321–330.

[706] Wurzelmann JI, Lyles CM, and Sandler RS. Childhood infections and the risk of inflammatory bowel disease. *Digestive Diseases and Sciences*, 1994;39(3):555–560.

[707] Xia L, Wang L, Chung AS, *et al.* Identification of both positive and negative domains within the epidermal growth factor receptor COOH-terminal region for signal transducer and activator of transcription (STAT) activation. *The Journal of biological chemistry*, 2002;277(34):30716–23.

[708] Xu J, Yang Y, Qiu G, *et al.* c-Maf regulates IL-10 expression during Th17 polarization. *Journal of immunology (Baltimore, Md. : 1950)*, 2009;182(10):6226–36.

[709] Xu WS, Parmigiani RB, and Marks PA. Histone deacetylase inhibitors: molecular mechanisms of action. *Oncogene*, 2007;26(37):5541–52.

[710] Yamazaki K, McGovern D, Ragoussis J, *et al.* Single nucleotide polymorphisms in TNFSF15 confer susceptibility to Crohn's disease. *Human molecular genetics*, 2005;14(22):3499–506.

[711] Yamazaki K, Umeno J, Takahashi A, *et al.* A genome-wide association study identifies 2 susceptibility loci for crohn's disease in a Japanese population. *Gastroenterology*, 2013;144(4):781–788.

[712] Yan R, Yang J, Jiang P, *et al.* Genetic variations in the SOCS3 gene in patients with Graves' ophthalmopathy. *Journal of clinical pathology*, 2015;68(6):448–52.

[713] Yang SK, Hong M, Zhao W, *et al.* Genome-wide association study of ulcerative colitis in Koreans suggests extensive overlapping of genetic susceptibility with Caucasians. *Inflammatory bowel diseases*, 2013;19(5):954–66.

[714] Yang SK, Hong M, Zhao W, *et al.* Genome-wide association study of Crohn's disease in Koreans revealed three new susceptibility loci and common attributes of genetic susceptibility across ethnic populations. *Gut*, 2014;63(1):80–7.

[715] Yang XJ and Seto E. HATs and HDACs: from structure, function and regulation to novel strategies for therapy and prevention. *Oncogene*, 2007;26(37):5310–8.

[716] Yantiss RK and Odze RD. Diagnostic difficulties in inflammatory bowel disease pathology. *Histopathology*, 2006;48(2):116–32.

[717] Yao Q, Shen M, McDonald C, *et al.* NOD2-associated autoinflammatory disease: a large cohort study. *Rheumatology*, 2015;54(10):1904–1912.

[718] Yao Q, Zhou L, Cusumano P, *et al.* A new category of autoinflammatory disease associated with NOD2 gene mutations. *Arthritis Research & Therapy*, 2011; 13(5):R148.

[719] Yarmus M, Woolf E, Bernstein Y, *et al.* Groucho/transducin-like Enhancer-of-split (TLE)-dependent and -independent transcriptional regulation by Runx3. *Proceedings of the National Academy of Sciences of the United States of America*, 2006;103(19):7384–9.

[720] Yin L and Chung WO. Epigenetic regulation of human β-defensin 2 and CC chemokine ligand 20 expression in gingival epithelial cells in response to oral bacteria. *Mucosal immunology*, 2011;4(4):409–19.

[721] Yoshikawa T, Takata A, Otsuka M, *et al.* Silencing of microRNA-122 enhances interferon-α signaling in the liver through regulating SOCS3 promoter methylation. *Scientific Reports*, 2012;2:1–10.

[722] Yoshimura A, Naka T, and Kubo M. SOCS proteins, cytokine signalling and immune regulation. *Nature reviews. Immunology*, 2007;7(6):454–465.

[723] Young MD, Wakefield MJ, Smyth GK, *et al.* Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome biology*, 2010;11(2):R14.

[724] Yun HS, Min YW, Chang DK, *et al.* Factors Associated with Vaccination among Inflammatory Bowel Disease Patients in Korea. *The Korean Journal of Gastroenterology*, 2013;61(4):203.

[725] Zahm AM, Thayu M, Hand NJ, *et al.* Circulating MicroRNA Is a Biomarker of Pediatric Crohn Disease. *Journal of Pediatric Gastroenterology and Nutrition*, 2011;53(1):26–33.

[726] Zanello G, Goethel A, Forster K, *et al.* Nod2 activates NF-kB in CD4+ T cells but its expression is dispensable for T cell-induced colitis. *PloS one*, 2013; 8(12):e82623.

[727] Zaru R, Ronkina N, Gaestel M, *et al.* The MAPK-activated kinase Rsk controls an acute Toll-like receptor signaling response in dendritic cells and is activated through two distinct pathways. *Nature immunology*, 2007;8(11):1227–1235.

[728] Zenewicz LA, Antov A, and Flavell RA. CD4 T-cell differentiation and inflammatory bowel disease. *Trends in molecular medicine*, 2009;15(5):199–207.

[729] Zeng Z, Feingold E, Wang X, *et al.* Genome-wide association study of primary dentition pit-and-fissure and smooth surface caries. *Caries research*, 2014; 48(4):330–8.

[730] Zhai Z, Wu F, Dong F, *et al.* Human autophagy gene ATG16L1 is post-transcriptionally regulated by MIR142-3p. *Autophagy*, 2014;10(3):468–79.

[731] Zhang L and Fang B. Mechanisms of resistance to TRAIL-induced apoptosis in cancer. *Cancer gene therapy*, 2005;12(3):228–237.

[732] Zhang L, Hou D, Chen X, *et al.* Exogenous plant MIR168a specifically targets mammalian LDLRAP1: evidence of cross-kingdom regulation by microRNA. *Cell Research*, 2012;22(1):273–274.

[733] Zhang Q, Pan Y, Yan R, *et al.* Commensal bacteria direct selective cargo sorting to promote symbiosis. *Nature immunology*, 2015;16(9):918–26.

[734] Zhao Y, Bjørbaek C, Weremowicz S, *et al.* RSK3 encodes a novel pp90rsk isoform with a unique N-terminal sequence: growth factor-stimulated kinase function and nuclear translocation. *Molecular and cellular biology*, 1995;15(8):4353–63.

[735] Zhernakova A, Festen EM, Franke L, *et al.* Genetic Analysis of Innate Immunity in Crohn ' s Disease and Ulcerative Colitis Identifies Two Susceptibility Loci Harboring CARD9 and IL18RAP. *Journal of Human Genetics*, 2008;(May):1202–1210.

[736] Zhou R, Hu G, Gong AY, *et al.* Binding of NF-kappaB p65 subunit to the promoter elements is involved in LPS-induced transactivation of miRNA genes in human biliary epithelial cells. *Nucleic acids research*, 2010;38(10):3222–32.

[737] Zhou X, Jiang Z, Zou Y, *et al.* Role of SOCS3 in the Jak/stat3 pathway in the human placenta: different mechanisms for preterm and term labor. *Acta Obstetricia et Gynecologica Scandinavica*, 2015;94(10):1112–1117.

[738] Zhuang J, Widschwendter M, and Teschendorff A. A comparison of feature selection and classification methods in DNA methylation studies using the Illumina Infinium platform. *BMC Bioinformatics*, 2012;13(1):59.

[739] Ziller MJ, Gu H, Müller F, *et al.* Charting a dynamic DNA methylation landscape of the human genome. *Nature*, 2013;500:477–81.

[740] Zimmermann N, Zschocke J, Perisic T, *et al.* Antidepressants inhibit DNA methyltransferase 1 through reducing G9a levels. *The Biochemical journal*, 2012; 448(1):93–102.

[741] Zuk O, Hechter E, Sunyaev SR, *et al.* The mystery of missing heritability: Genetic interactions create phantom heritability. *Proceedings of the National Academy of Sciences of the United States of America*, 2012;109(4):1193–8.